

INFORMATIONSPOTENTIALE. VOM KOMMUNIZIEREN MIT DIGITALEN DATENBANKEN

In *Grammophon, Film, Typewriter* hat Friedrich Kittler 1986 das Ende der Medien ausgerufen, welches sich im Aufkommen des Computers ankündigt. Durch den Computer als Medium kommen *die* Medien an ihr Ende. Dieses Ende ist jedoch ein doppeltes, es bedeutet den Tod und die Vollendung der Medien:

»In der allgemeinen Digitalisierung von Nachrichten und Kanälen verschwinden die Unterschiede zwischen einzelnen Medien. Nur noch als Oberflächeneffekt, wie er unterm schönen Namen Interface bei Konsumenten ankommt, gibt es Ton und Bild, Stimme und Text. Blendwerk werden die Sinne und der Sinn. Ihr Glamour, wie Medien ihn erzeugt haben, überdauert für eine Zwischenzeit als Abfallprodukt strategischer Programme. In den Computern selber dagegen ist alles Zahl: bild-, ton- und wortlose Quantität« (Kittler 1986, 7f.).

Das Verschwinden der Einzelmedien ist Resultat der Vollendung der Medien im Universalmedium Computer, welches sich gerade dadurch auszeichnet, alle vormaligen Medien integrieren zu können.◀1 Gemeinsame Basis aller Medien bildet fortan der digitale Code, auf dessen Grundlage es Kittler zufolge möglich ist »jedes Medium in jedes andere Medium« (Kittler 1986, 8) zu übersetzen. An die Stelle der Differenz der Medien tritt, so die Überzeugung Kittlers, die Einheit der Information, welche in der nachrichtentechnischen Kommunikationstheorie durch Claude E. Shannon (1948) auf ihr mathematisches Fundament gestellt wurde.◀2

Seit der Publikation von *Grammophon, Film, Typewriter* ist ein Vierteljahrhundert vergangen. In der Zwischenzeit sind Computer und mit ihnen das Internet zu einem alltäglichen und deshalb zentralen Bestandteil der globalen Medienkultur geworden, was jedoch zur Folge hatte, »dass zum Ärger technischer Puristen, die Befehlsausführung auf Ebene des Betriebssystems zunehmend hinter der ikonischen Ebene des Displays verschwindet« (Hartmann 2006, 194). *Sinn* und *Sinnlichkeit* erwiesen sich entgegen Kittlers Prognose also nicht als temporäres *Blendwerk*, sondern haben auch unter den Bedingungen der fortschreitenden Digitalisierung Bestand. Der Computer als Medium nivelliert nicht sämtliche Mediendifferenzen, sondern fungiert als Metamedium auf dessen Bühne die Unterschiede zwischen Medien als Spiel medialer Formen bis zum

Exzess getrieben werden (vgl. Laurel 1993). Hierin mag man einen Missbrauch des Computers erkennen, der dessen eigentliche Bestimmung unterwandert, »das absolute Wissen als Endlosschleife« (Kittler 1986, 8) zu prozessieren. Oder man beginnt den Computer von seinen kontingenten Gebrauchsformen her als eine Medientechnologie zu betrachten, die unsere kommunikativen und informationellen Praxen auf mannigfaltige und zum Teil auch widersprüchliche Weise konfiguriert (und immer wieder auch rekonfiguriert). Eine solche Betrachtung geht nicht von der Programmierbarkeit des Computers aus, deren Prinzip theoretisch im Modell der Turingmaschine (1987 [1937]) beschrieben und praktisch in der Von-Neumann-Architektur (1987 [1945]) umgesetzt wurde, sondern analysiert Computer als programmierte Technologien, deren Medialität sich vor dem Hintergrund ihrer jeweils spezifischen Programme abzeichnet. Information wird infolgedessen nicht als »sub-semantic Effekt der Hardware von Kommunikation« (Ernst 2002, 173) verstanden, sondern als ein Effekt programmierter Computerhardware, das heißt von Software-Hardware-Konfigurationen.◀3

Computer prozessieren Medienobjekte (Texte, Bilder, Filme, Animationen, Messdaten) nicht (nur)◀4 unterschiedslos als Information im nachrichtentechnischen Sinn. Was Information ist, zeigt sich erst im Kontext der programmgesteuerten Verarbeitung digital codierter medialer Konstellationen; denn Software bedingt mit, was als Information zur Erscheinung kommen, als solche gespeichert, übertragen und verarbeitet werden kann. An die Stelle des Mythos eines einheitlichen und universellen Informationskonzepts tritt dabei die tatsächliche Vielfalt partikularer Formen von Information, wie sie durch konkrete Computeranwendungen hervorgebracht werden.◀5 Hiervon zeugen die konzeptuelle und technische Entwicklung digitaler Datenbanken sowie die medialen Praxen, in die Datenbanktechnologien in unserer zeitgenössischen Medienkultur verflochten sind. Dies gilt es im Folgenden nachzuzeichnen.

Ausgehend von einer Rekonstruktion des doppelten Ursprungs digitaler Datenbanken im Managementdiskurs und im Diskurs der bibliothekarischen Informationsverwaltung werden zwei Modelle der Kommunikation mit Informationssammlungen skizziert, die nicht nur ein Gegenmodell zu Shannons nachrichtentechnischem Kommunikationsmodell darstellen, sondern auch einen spezifischen Problemkontext beschreiben, der den Horizont der Entwicklung konkreter Datenbanktechnologien bildet. Vor diesem Hintergrund wird deutlich, dass die computertechnische Verarbeitung von Informationen keineswegs grundsätzlich von der Frage nach deren Sinn, Gehalt und Bedeutung absieht. Gleichwohl Computer Informationen nur auf dem Niveau ihrer formalen Merkmale regelgeleitet verarbeiten können, folgt hieraus nicht notwendig ein

Verzicht auf Semantik. Im Gegenteil, die Entwicklung digitaler Datenbanken kann als Experiment verstanden werden, Semantik in Syntax zu übersetzen und hierdurch die Verarbeitung von Semantik in nicht-semantischen Routinen zu ermöglichen.

Im zweiten Schritt wird die Übersetzung der abstrakten Modelle der Datenbankkommunikation in konkrete technische Probleme und Lösungsstrategien dargestellt. Dabei wird die These verfolgt, dass die Einführung der Festplatten hardwaretechnischen Horizont bildet, vor dem sich die Entwicklung von zentralen Konzepten der Informationsverwaltung in digitalen Datenbanken sowie die technische Implementierung von Datenbanksystemen vollzieht. Dies wird an der Entwicklung hin zur ANSI/X3/SPARC-Datenbankarchitektur veranschaulicht, welche noch immer als Modell der Informationsverwaltung in digitalen Datenbanken dient. Das prinzipielle Ziel, die Findbarkeit von Informationen in Datenbanken zu gewährleisten, wird in dem Entwurf dieser Datenbankarchitektur von dem ökonomisch motivierten Wunsch begleitet, die Unabhängigkeit digitaler Informationen von den sie verarbeitenden Programmen sicherzustellen. An diese technische Problemkonstellation anschließend, kann der Debatte um Materialität respektive Immaterialität digitaler Information eine produktive Wendung gegeben werden, indem gezeigt wird, dass die Immaterialität digitaler Informationen weder ein bloßer Schein ist, der durch die irreduzible Materialität von Computern und Computernetzwerken entlarvt wird, noch dass es sich um ein Wesensmerkmal dieser Informationen handelt. ◀6 Vielmehr ist der Eindruck der Immaterialität ein Effekt der zunehmenden Entkopplung der Informationsverarbeitung von ihrer Verwaltung im Computer, die durch Datenbanken gewährleistet wird und sich als Autonomie digitaler Informationen gegenüber den sie verarbeitenden Programmen respektive Präsentations-, Zugriffs- und Auswertungsformen zeigt.

Kommunikation mit Datenbanken

Die digitale Medienkultur ist nicht nur eine Kultur der Algorithmen, der Berechnung, der Simulation und der Generierung digitaler Welten, sie ist gleichermaßen eine Kultur des Sammelns, Speicherns, Zirkulierens und Suchens von Informationen. Ende der 1950er und Anfang der 1960er Jahre wurde die Idee der computergestützten Versammlung, Speicherung und Abfrage von Informationen auf den Begriff der Datenbank gebracht. Bezeichnet wird hierdurch ein Speicher aus dem im Zugriff Informationen geschöpft werden können. Im Zuge der Herausbildung und Popularisierung der Datenbankidee fand im Juni 1963

in Santa Monica, Kalifornien, die wohl erste Tagung statt, welche sich explizit und ausschließlich dem Problem der Versammlung und Verwaltung von Informationen in Computern widmete. Auf Initiative der *System Development Corporation* (SDC) und der *Advanced Research Projects Agency* (ARPA), deren *Information Processing Techniques Office* zu dieser Zeit von J.C.R. Licklider geleitet wurde, trafen bei dem Symposium »Development and Management of a Computer-Centered Data Base« Entwickler und potenzielle Nutzer aufeinander, um über die möglichen Anwendungsgebiete von Datenbanken zu diskutieren und um sich über konzeptuelle und technische Grundsatzfragen zu verständigen. Wie Licklider in seinen Eröffnungsbemerkungen feststellt, war eine der vordringlichsten Fragen, die dabei im Raum standen, die Klärung des Begriffs der Datenbank: »Early in the conference, someone should define ›data base« (Licklider 1964, 1). Am Ende der Tagung stand entgegen der Hoffnung Lickliders nicht eine Definition von Datenbanken, sondern zwei. Dies ist nicht zuletzt auch deshalb bedeutsam, weil in der Selbstgeschichtsschreibung der Informatik sowie in den kulturwissenschaftlichen Hinwendungen zur Geschichte digitaler Datenbanken gemeinhin nur auf einen der beiden Definitionsvorschläge rekuriert wird, welcher Datenbanken anhand von drei Kriterien bestimmt: »1. A database is a set of files. 2. A file is an ordered collection of entries. 3. An entry consists of a key, or keys, and data« (Franks 1964, 119).⁴⁷ Dieses Verständnis von Datenbanken als einer Menge von Dateien, in denen Einträge versammelt sind, die Daten beinhalten und durch Schlüssel adressiert werden können, greift auf ein Vokabular zurück, das sich bereits in den 1930er Jahren beim Umgang mit Lochkartenmaschinen im Bereich der technischen Daten- und Informationsverarbeitung etabliert hat (vgl. Haigh 2009, 7). Medienhistorisch lässt sich die Unterscheidung von Dateien, Einträgen, Daten und Schlüssel jedoch noch weiter zurückverfolgen, wie Cornelia Vismann in ihrer Studie zu Akten gezeigt hat:

»Das US-Militär wird Jahrhunderte später den Begriff ›Datenbank‹ medienhistorisch präzise so definieren, dass darin die diachrone Entwicklung von Akten aus Registern zu Indices enthalten ist. Aus Akten (*files*) werden Einträge / Regesten (*entries*) deduziert, die wiederum eine Doppelfunktion als Schlüssel und Daten (*key and data*) übernehmen« (Vismann 2001, 170f.).

Konzeptuell werden Datenbanken hierdurch in den Kontext der administrativen und institutionellen Informationsverarbeitung gestellt, der auch praktisch im Vordergrund des Datenbanksymposiums stand. Die zentrale Frage war, wie Computerdatenbanken zur Verbesserung des Managements großer Organisationen – seien es wirtschaftliche Unternehmen oder Armeen – beitragen können. Deutlich wird dies unter anderem in dem von Colonel A. K. Swanson vorgebrachten Problem, der organisationsinternen Auswahl von ge-

eigneten Kandidaten für bestimmte Aufgaben, welches er am Beispiel der Abordnung von Offizieren in der *United States Air Force* diskutierte (vgl. Swanson 1964, 54f.). Die Datenbank gewinnt hierbei Kontur als eine Technologie des *Controlling*, die die Überwachung und Steuerung von Organisationen unterstützen soll.

Anders als es die geläufige Zitationspraxis suggeriert, erzielte man während der Tagung noch keine Einigung darüber, was genau Datenbanken sind beziehungsweise welche Aufgabe sie erfüllen sollen. Wie E. W. Franks berichtet, wurde an der bereits angeführten Definition auch Kritik laut, die auf den relativ engen Fokus des Definitionsvorschlags abzielt. In Reaktion hierauf erwogte man eine zweite breitere, aber auch unspezifischere Definition, der zufolge eine Datenbank erstens aus Daten und zweitens aus Mitteln zum Zugriff auf diese Daten besteht (vgl. Franks 1964, 120).⁴⁸ Dieses weite Verständnis weist in Richtung eines zweiten Entwicklungskontexts, aus dem heraus die medientechnische Praxis digitaler Datenbanken erwachsen ist. So lässt sich die konzeptuelle und technische Entwicklung von digitalen Datenbanken nicht nur auf den Bereich der institutionellen Informationsverarbeitung zurückführen, vielmehr steht die Datenbankentwicklung auch in der Tradition des bibliothekarischen Diskurses des *Information Storage and Retrieval*.

Geprägt wurde der Begriff des *Information Retrieval* 1950 von Calvin Mooers, dessen Problemhorizont er folgendermaßen definiert: »The problem of directing a user to stored information, some of which may be unknown to him, is the problem of ›information retrieval‹« (Mooers 1950, 572). Vor diesem Hintergrund formuliert Mooers ein Modell der technischen Kommunikation mit Informationssammlungen:

»In information retrieval, the addressee or receiver rather than the sender is the active party. Other differences are that communication is temporal from one epoch to a later epoch in time, though possibly at the same point in space; communication is in all cases unidirectional; the sender cannot know the particular message that will be of later use to the receiver and must send all possible messages; the message is digitally representable; a ›channel‹ is the physical document left in storage which contains the message; and there is no channel noise because all messages are presumed to be completely accessible to the receiver« (Mooers 1950, 572).

Gleichwohl Mooers das nachrichtentechnische Kommunikationsmodell nicht explizit erwähnt, buchstabiert er im Detail die Gemeinsamkeiten und Unterschiede zwischen den beiden Formen von Kommunikation aus, die er einerseits als Kommunikation im Raum und andererseits als Kommunikation in der Zeit charakterisiert.⁴⁹

Während Shannon Kommunikation vorrangig vom Problem der Überwindung räumlicher Distanzen her als »point-to-point signalling« (Mooers 1950, 572) denkt, betrachtet Mooers die Kommunikation des *Information Retrieval* als »temporal signalling« (Mooers 1950, 572). Dies hat erstens eine Umkehr der Aktivitätszuschreibung zur Folge: Nicht der Sender, sondern der Empfänger ist der aktive Part der Kommunikation. Zweitens ist die Kommunikation strikt unidirektional, weshalb es notwendig wird, dass nicht nur bestimmte Nachrichten ausgesandt werden, sondern alle möglichen Nachrichten. Schließlich geht Mooers drittens davon aus, dass es das Problem des Kanalauschens im Kontext des *Information Retrieval* nicht gibt. Auf Seiten des Empfängers liegen die Informationen, so Mooers Annahme, in Form von Dokumenten vollständig, intakt und demzufolge rauschfrei vor. Diese Behauptung lässt sich als eine strategische Idealisierung verstehen, die es Mooers erlaubt, Kommunikation als den suchenden Zugriff auf einen Speicher zu modellieren. Im Vordergrund steht dabei eben nicht die Übertragung von Nachrichten zwischen Personen, sondern die mehr oder minder zielgerichtete Abfrage von Informationen, die gespeichert vorliegen. Damit sind die Eckpfeiler eines Kommunikationsmodells skizziert, welches ebenso wie es auch bei Shannons Modell der Fall ist, nicht Kommunikation im Allgemeinen, sondern ein spezifisches Kommunikationsproblem beschreibt, das im Finden von Informationen besteht. ◀10

Die einseitige Fokussierung auf das (Auf)Finden von Dokumenten führte dazu, dass Mooers der Seite der Bereitstellung von Informationsressourcen und dem Speicher noch keine besondere Aufmerksamkeit beimaß. Dies änderte sich Ende der 1950er Jahre, als das Kommunikationsmodell des *Information Retrieval* in einem dezidierten Modell der Datenbankkommunikation kulminierte, welches im Rahmen eines vom *Office of Naval Research* finanzierten Forschungsprojekts bei der *Benson-Lehner Corporation* entwickelt wurde. »Data Banking«, so ist im Abschlussbericht des Projekts zu lesen, »is the process of communicating between many conceivers to many receptors through a store« (Worsley et al. 1959, 5f.). Der Begriff des *Data Banking* dient den Autoren als Oberbegriff, um den gesamten Prozess des *Information Storage* und des *Information Retrieval* zu beschreiben. Es geht dabei nicht mehr nur um die Kommunikation mit Informationssammlungen, sondern um eine Kommunikation durch Datenbanken, die als fünffacher Transformationsprozess modelliert wird.

Gedanken werden in Dokumente übersetzt (1), die in einen Dokumentspeicher eingehen (2). An diesen Speicher können Nutzer Fragen richten, die sie auf Grundlage ihres Informationsbedürfnisses formulieren (3). In Reaktion hierauf werden aus dem Speicher diejenigen Dokumente ausgewählt, die der Anfrage entsprechen (4) und an den Nutzer übermittelt, der diese im letzten

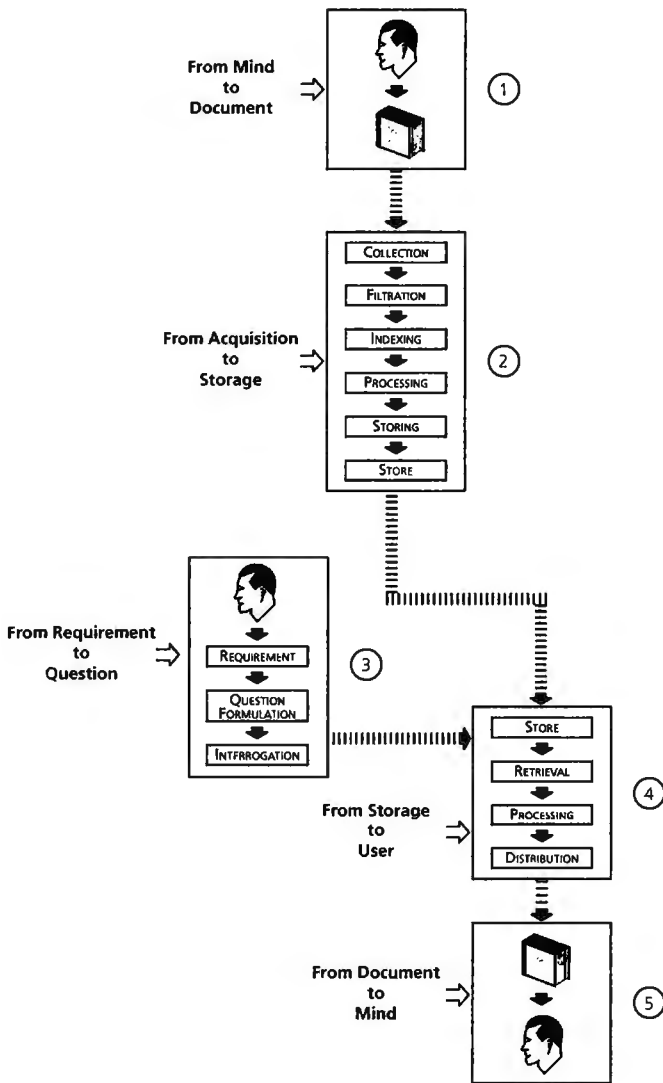


Abb. 1: Modell des Data Banking ◀11

Schritt in Verständnis übersetzt (5). Zwischen den Produzenten von Nachrichten und ihren Rezipienten vermittelt diesem Modell zufolge kein Kanal, sondern ein *Store*, den die Autoren im folgenden *Data Bank* nennen. Als Technik sollen Datenbanken die in Punkt zwei und vier spezifizierten Funktionen der

Sammlung, Filterung, Indexierung, Verarbeitung, Speicherung, Abfrage und Distribution erfüllen. Eine Datenbank beziehungsweise ein Datenbanksystem ist demzufolge kein passiver Container, der Informationen nur aufbewahrt, vielmehr werden Informationen in Datenbanken zu einem Bestand, über den die Nutzer verfügen können. Die Datenbank birgt ein Informationspotenzial, welches im Prozess der Abfrage aktualisiert werden muss. Im suchenden Zugriff auf ein Reservoir vorhandener Informationen kommuniziert der Nutzer, indem er bestimmte Informationen selektiert und durch diese überrascht wird.¹¹² Hierdurch erfährt nicht zuletzt auch das Konzept der Information eine grundlegende Transformation: Wird Information in der Nachrichtentechnik daran bemessen, welche Selektionen der Äußerung einer Nachricht inhärent sind, widmen sich Datenbankdiskurs und Datenbankpraxis auf der anderen Seite der Herausforderung aus einer Zahl von Nachrichten respektive aus einem Reservoir vorhandener Informationen diejenigen auszuwählen, die das nutzerseitige Informationsbedürfnis befriedigen. Kurzum: Der Gegenstand von Shannons Analysen ist die Gesamtheit möglicher Nachrichten und Äußerungen. Demgegenüber richtet sich das *Information Storage and Retrieval* nur auf getätigte Äußerungen und Aussagen: »Library communication differs from telegraphic mainly in that all the messages have been sent already, and you have to pick out the right one to suit a query not known beforehand« (Fairthorne 1961b [1953], 25). Über Informationen verfügen zu können, bedeutet im Kontext der Kommunikationsmodelle daher auch unterschiedliches. Shannon ist mit dem Problem der Replikation von Nachrichten beschäftigt, weshalb der Empfänger im Rahmen seines Modells genau dann über eine Information verfügt, wenn er die ausgesandte Nachricht vollständig empfangen hat. Dies wird im Modell der Datenbankkommunikation als unproblematisch vorausgesetzt. Die Verfügbarkeit von Information bedeutet hier nicht nur das Besitzen oder Haben von Nachrichten beziehungsweise Informationen, sondern gründet auf der Möglichkeit, diese aus einem Bestand vorhandener Informationen auszuwählen zu können.

Infolgedessen tritt genau *der* Aspekt von Information in den Vordergrund theoretischer Betrachtungen und praktischer Erwägungen, den Shannon ausklammern konnte: Die Bedeutung von Information im doppelten Wortsinn von Gehalt und Relevanz. Dass Nachrichten auch eine Bedeutung haben, kann vor dem Hintergrund des Problems der nachrichtentechnischen Übertragung von Signalen ignoriert werden. »Frequently«, so schreibt Shannon, »the messages have *meaning*; that is they refer to or are correlated to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem« (Shannon 1948, 379).

Bedeutung ist für den Nachrichtentechniker aus zwei Gründen bedeutungslos. Vor dem Hintergrund des Übertragungsproblems ist es erstens geboten von der Bedeutung von Nachrichten abzusehen, da der Kanal nicht nur besondere Nachrichten, sondern jede mögliche Nachricht übertragen können soll: »The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design« (Shannon 1948, 379). Zweitens werden nicht Bedeutungen übertragen, sondern Nachrichten. Dass diese für den Sender und den Empfänger einen Sinn haben, ist dabei nebensächlich. Aufgrund dessen war es für Shannon ein Leichtes, die Frage nach dem Informationsmaß für Nachrichten beziehungsweise der Übertragungskapazität eines Kanals von der Frage nach der Bedeutung von Information zu entkoppeln.

Im Bereich der Datenbankkommunikation ist die Frage des (computer)technischen Umgangs mit Bedeutung jedoch entscheidend. Dies zieht ein grundlegendes Problem nach sich, denn Computer sind nicht in der Lage Bedeutung zu verstehen, das heißt sie interpretieren Information nicht semantisch, sondern verarbeiten sie auf dem Niveau ihrer Syntax. Können Computer nicht verstehen, was sie sammeln, indexieren und wieder zugänglich machen sollen, dann scheint das Ziel digitaler Datenbanken von vornherein zum Scheitern verurteilt und Shannons theoretisch motivierte Einklammerung von Bedeutung praktisch bestätigt. Hierbei handelt es sich jedoch um einen Kurzschluss, der auf der falschen Annahme beruht, dass der Umgang mit semantischen Informationen, das heißt mit Sinn, Gehalt und Bedeutung, nur auf der Grundlage von Verstehen und Verständnis möglich ist. Robert Fairthorne zieht eben diese Vorstellung in Zweifel, wenn er feststellt, dass der bibliothekarische Umgang mit Dokumenten keinesfalls immer auf hermeneutischer Interpretation und Verständnis beruht. Im Gegenteil, viele Tätigkeiten im Kontext der Erschließung, Verwahrung und Verfügbarmachung von Dokumenten sind ein regelbasierter und routinemäßiger Umgang mit Dokumenten und nicht das Resultat von verstehender Interpretation. Hierbei handelt es sich Fairthorne zufolge um »clerical activities« (Fairthorne 1961d, 94), die ebenso gut von Menschen wie von Computern erfüllt werden können. Was Fairthorne beschreibt, ist eine Ebene der nichtsemantischen Verarbeitung von Semantik. Auf dieser Ebene ist die Bedeutung von Informationen durch die Regeln des Umgangs mit diesen Informationen definiert: »For very many purposes the ›meaning‹ of a set of symbols is adequately defined as the rules for its use« (Fairthorne 1961c [1954], 66). Die Formulierung legt bereits nahe, dass Fairthorne keineswegs davon ausgeht, Computer könnten Informationen so wie Menschen verstehen oder dass das menschliche Verstandes- und Verständnisvermögen medientechnisch obsolet

würde. Vielmehr geht es ihm darum den Bereich auszuloten, in dem Computer erfolgreich zur Verwaltung und Verarbeitung bedeutungsvoller Informationen eingesetzt werden können. Hierfür sei es notwendig Konzepte in Physik zu übersetzen, das heißt Information über Realität in Information als Realität. Die in Dokumenten implizit enthaltenen semantischen Informationen sollen, so sein Vorschlag, durch Markierungen am Dokument expliziert werden: »The bridge between the concepts and the physics of retrieval is the notation or system of marking the text« (Fairthorne 1961a, X). Neben der Markierung, verstanden als Hinzufügung von Inskriptionen zu Dokumenten, zieht Fairthorne ein zweites Verfahren der Explikation des Impliziten in Betracht – die Einordnung der Dokumente in ein Ordnungssystem. Bedeutung wird dabei nicht durch ein materielles Zeichen am Dokument manifest, sondern lässt sich aus der Stelle herauslesen, die ein Dokument in einem Ordnungssystem, wie zum Beispiel der Dewey-Dezimalklassifikation, einnimmt.◀13 Durch »marking« and »parking.« (Fairthorne 1961b [1953], 95) können in Dokumenten enthaltene semantische Informationen mittelbar durch Technologien in routinemäßigen Verfahren gehandhabt werden. Dass der computertechnische Umgang mit semantischen Informationen im Vergleich zum menschlichen Interpretationsvermögen defizitär ist, versteht sich dabei einerseits von selbst, ist andererseits aber nicht relevant, da der menschliche Gebrauch von Computern zur Verwaltung und Verarbeitung von Semantik im Zentrum des Interesses steht und nicht der Versuch, den Menschen durch Computer zu ersetzen. Infolgedessen wird das semantische Wissen menschlicher Interpreten nicht überflüssig. Das Aufstellen der Übersetzungs- und Verarbeitungsregeln setzt Fairthorne zufolge Verstehen und Verständnis voraus. Sind diese Regeln jedoch festgelegt, kann Semantik auch in nichtsemantischen Prozeduren verarbeitet werden. Grundlegend hieran ist die Einsicht, dass die (computer)technische Verarbeitung von Information nicht auf einer systematischen Ein- beziehungsweise Ausklammerung von Semantik beruhen muss. Lassen sich semantische Informationen in gewissen Grenzen in syntaktische Informationen übersetzen, dann ist das Absehen von Semantik kein Effekt der Hardware, sondern der Effekt einer theoretischen Prädisposition, die sich von dem zugrunde gelegten nachrichtentechnischen Kommunikationsmodell herleitet. Die beschriebenen Modelle der Kommunikation mit Informationssammlungen bilden hierzu eine Alternative. Vor ihrem Hintergrund bedeutet die algorithmische Verarbeitung von Informationen im Computer keinen Verzicht auf oder Abschied von Semantik, sondern lässt sich als Modus des nicht-interpretativen operativen Umgangs mit Zeichen verstehen, wodurch Sybille Krämer zufolge ein zentrales Charakteristikum von Kul-

turtechniken beschrieben ist (vgl. 2003, 169f.). In Anbetracht dessen lassen sich Datenbanktechnologien als Kulturtechnologien begreifen.

Vom Direct Access zur ANSI/X3/SPARC-Datenbankarchitektur

Artikuliert wurden in den Modellen der Kommunikation mit Informationssammlungen abstrakte Ziele für beziehungsweise Aufgaben von Datenbanken. Im Zuge der Realisierung digitaler Datenbanktechnologien mussten diese in konkrete technische Probleme und ingenieurtechnische Lösungsstrategien übersetzt werden. Die Versammlung und Abfrage von Informationen mit Computern erwies sich auf der grundlegenden Ebene der Speicher dabei als ein Adressierungsproblem. Auch auf digitalen Datenträgern haben Informationen einen physischen Ort: als materielle Inskriptionen nehmen sie Raum ein. Das Finden einer bestimmten Information, wie zum Beispiel die Höhe der Neuverschuldung Deutschlands im Jahr 2010, bedeutet den Ort dieser Information aufzusuchen oder genauer, die an einem bestimmten Ort befindlichen Inskriptionen auszulesen. Ist der Ort der Information unbekannt, dann kann man nach dieser nur im gesamten Speicherraum suchen. Für den Fall, dass die Anordnung der Informationen im Speicher keiner systematischen Ordnung unterliegt, käme dies der sprichwörtlichen Suche nach der Nadel im Heuhaufen und damit dem Verlust der Information gleich. Analog hierzu gelten Bücher, die in Bibliotheken nicht entsprechend der vorgesehenen Stellordnung einsortiert werden, über kurz oder lang als verloren. Der Vergleich mit Bibliotheken ist in einer weiteren Hinsicht instruktiv. Die Anordnung von Medienobjekten im Raum der Bibliothek wird hier überlagert von einer zweiten Ordnung, der Katalogordnung. Letztere konstituiert ein Ordnungssystem des Wissens, das das Finden von Büchern in der Bibliothek ermöglichen beziehungsweise vereinfachen soll. Doch die Verzeichnung von Büchern im Katalog entlastet den Bibliotheksbesucher nicht von der Notwendigkeit sich auch in der Raumordnung der Bibliothek zu orientieren und durch diese zu navigieren. Dies aber müssen digitale Datenbanken leisten, die das Vergessen des Orts von Daten im Speicher ermöglichen sollen, indem sie es erlauben, Daten als Informationen zu adressieren. Das Herausfinden, welche Informationen es in der Datenbank gibt und das Auffinden dieser Informationen fallen dabei in eins.

Als *Information Retrieval* Technologien dienen Datenbanken zur Automatisierung der Suche. Nutzerseitige Anfragen werden in Suchroutinen übersetzt, die in der Datenbank automatisch diejenigen Informationen ausfindig machen

sollen, nach denen der Nutzer fragt. ◀14 Eine mögliche Strategie, dies technisch umzusetzen, ist das schrittweise Durchsuchen des gesamten Speichers. So lange Speichertechnologien, wie Lochkarten oder Magnetbänder, nur den seriellen Zugriff auf den Gesamtbestand an gespeicherten Informationen erlaubten, war kein anderes Vorgehen bei der automatisierten Suche durch Computer denkbar. Die Beschränkung auf seriellen Zugriff wirkte sich dabei nicht nur auf die Abfrage von Information aus, sondern auch auf ihre Speicherung. Zusammengehörige Informationen wurden in seriellen Speichern als Datensätze angeordnet, die wiederum geordnet nach einem bestimmten Schlüsselwert auf dem Datenträger abgelegt wurden. ◀15 Hierdurch versuchte man die Nachteile des seriellen Zugriffs bei der Abfrage von Information zu kompensieren.

Erst mit der Markteinführung der Festplatte 1956 durch IBM standen alternative Strategien der automatisierten Suche nach Informationen in Datenbanken zur Disposition. Bei der Festplatte handelt es sich um eine Technologie, die nicht nur die dauerhafte Speicherung relativ großer Mengen digitaler Information erlaubt, sondern vor allem auch den Direktzugriff auf diese ermöglicht. Der entscheidende Vorteil von Festplatten im Vergleich zu den anderen in den 1950er Jahren verfügbaren Massenspeichertechnologien war, dass jeder Ort im Speicher in einer konstant kurzen Zeit erreicht werden kann (vgl. Lesser/Haanstra 1957, 140). Die durchschnittliche Zugriffszeit auf einen beliebigen Speicherort reduziert sich von fünf Minuten auf eine sechstel Sekunde (Bachman 1962a). Dieser Performancegewinn, den die Einführung von Festplattenspeichern mit sich brachte, bedeutete keineswegs die Lösung aller Probleme der computertechnischen Informationsverarbeitung. Vielmehr traten im Zuge der Einführung von Festplatten die technischen Herausforderungen der Versammlung und Abfrage von Information als Adressierungsproblem erst richtig zum Vorschein. Schnell zugegriffen werden kann nämlich nur auf diejenigen Informationen, deren Ort im Speicher bekannt ist. ◀16 Geht es aber darum in einer Datenbank neue Informationen zu finden, dann ist die Adresse der gesuchten Information ja gerade unbekannt.

Wie Charles Bachman Anfang der 1960er Jahre darlegt, ist die von Festplatten eröffnete Möglichkeit, jede beliebige Information schnell abzufragen, mit Hilfe von Datenbanktechnologien in das Vermögen zu übersetzen, ebenso schnell spezifische Informationen finden zu können (vgl. Bachman 1962b). Das nutzerseitige Informationsbedürfnis gelte es mit den im Speicher ruhenden Informationen effizient zu koordinieren. Formalisiert wurde diese Übersetzungsaufgabe Ende der 1960er Jahre in den Berichten der *CODASYL Data Base Task Group* (1969, 1971) als Vermittlung zwischen verschiedenen Informationsmodellen, die im Entwurf von Datenbanken spezifiziert werden müssen (Abb.2).

Das Schema dient auf der einen Seite der Definition des Informationsmodells, mit dem der Computer operiert und auf Grundlage dessen die Informationen im Speicher abgelegt werden. Ein Subschema spezifiziert auf der anderen Seite die Nutzersicht auf die Datenbank, welche als eigenes Informationsmodell konzeptualisiert wird. Obgleich Schema und Subschema miteinander kompatibel sein müssen, wird durch die Unterscheidung der beiden Ebenen der Tatsache Rechnung getragen, dass Computer und Nutzer auf unterschiedliche Weise mit Information operieren. Darüber hinaus kommt in der Unterscheidung der beiden Beschreibungsebenen ein weiteres wichtiges Motiv zum Vorschein, welches die Entwicklung von Datenbanktechnologien anleitete. Würden keine Nutzersichten auf Information definiert, dann hätte dies zur Konsequenz, dass sich jede Änderung an der internen Anordnung von Informationen im Speicher und gegebenenfalls jeder Wechsel der Computerhardware auf die Benutzerschnittstellen zur Datenbank und damit auf die Datenbanknutzer auswirkt. Insofern dient die Einführung von Subschemata der Abschirmung von Nutzern respektive Anwendungsprogrammen gegenüber Änderungen an der Datenbank in der unsichtbaren Tiefe des Computers (vgl. Luhmann 1998, 302ff.). Gerade weil die Entwicklung von Computeranwendungen zu dieser Zeit sehr aufwendig war und damit einen wichtigen Kostenfaktor darstellte, sollte die Unabhängigkeit von Anwendungsprogrammen gegenüber Änderungen in der Datenbank gewährleistet werden. Dieses Ziel wird seit Ende 1960er Jahre unter dem Stichwort *Datenunabhängigkeit* diskutiert und ist seither eines der zentralen Leistungsmerkmale von Datenbanksystemen.

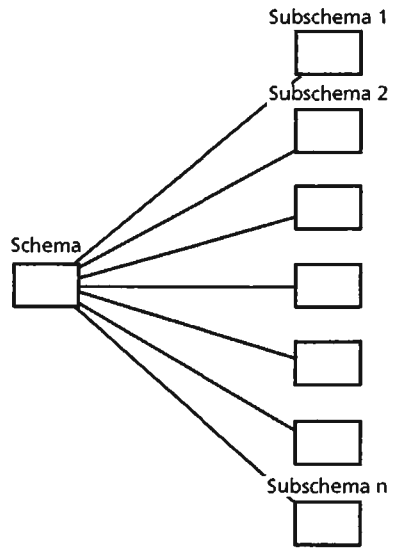


Abb.2: CODASYL Schema-Subschema-Architektur

Auch wenn die Mitglieder der *CODASYL Data Base Task Group* zunächst glaubten, mit der Unterscheidung von Schema und Subschema einen Vorschlag unterbreitet zu haben, der bis zu einem gewissen Grad Datenunabhängigkeit gewährleistet (1971, 18), hat sich dies in den Folgejahren als nicht hinreichend erwiesen. Im Rahmen der Standardisierungsbemühungen der *ANSI/X3/SPARC Study Group on Database Systems* wurde die Zwei-Ebenen-Architektur daher zu einer Drei-Ebenen-Architektur von Datenbanken erweitert (1975). Zwar war auch die-

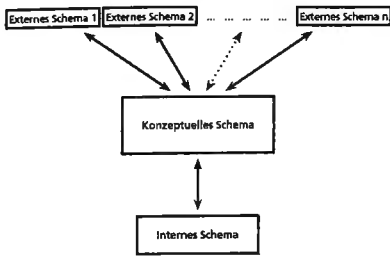


Abb.3: ANSI/X3/SPARC-Datenbankarchitektur

se Datenbankarchitektur nicht geeignet, um die Implementierung von Datenbankmanagementsystemen praktisch anzuleiten, konzeptuell aber bildet sie heute noch immer den Rahmen, in dem sich der Entwurf von und der Umgang mit Datenbanken vollzieht. Der Unterscheidung von Schema und Subschema wird eine weitere hinzugefügt, die das Modell der internen Repräsentation von Informationen im Speicher von seiner konzeptuellen Beschreibung entkoppelt. Infolgedessen werden systematisch drei Ebenen des Umgangs mit respektive der Modellierung

von Information in Datenbanken unterschieden, die als internes, konzeptuelles und externes Schema bezeichnet werden (Abb. 3). Das konzeptuelle Schema bildet dabei das Zentrum und die Scharnierstelle zwischen dem Computer und dessen Eigenlogik einerseits und seinen Nutzern und deren Gebrauchslagen andererseits. In diesem wird formal deklariert, welche Informationen in einer Datenbank versammelt und aus dieser abgefragt werden können. «17 Als Mittelglied zwischen der internen Ordnung von Information im Speicher und den externen Sichten auf die Datenbank, ermöglicht das konzeptuelle Schema deren wechselseitige Unabhängigkeit: »The placement of the conceptual schema between an external schema and the internal schema is necessary to provide the level of indirection essential to data independence« (Tschritzis/ Klug 1978, 184). Effekt hiervon sind die scheinbar grenzenlosen Möglichkeiten des Umgangs mit Datenbankinformationen auf den Benutzeroberflächen verschiedener Programme und die Freiheit zum Entwurf immer neuer *User Interfaces* zu Datenbanken. Diese Möglichkeit manifestiert sich in der heutigen Medienkultur auf exemplarische Weise im Phänomen des *Mashup* – das heißt in der Kombination von Informationen aus verschiedenen Quellen, die über *Application Interfaces* abgefragt werden – zu einem neuartigen Medienobjekt beziehungsweise einer neuartigen Anwendung.

Angesichts der mannigfaltigen Möglichkeiten mit Informationen auf den Benutzeroberflächen heutiger Computer umzugehen, sie zu selektieren, zu präsentieren, auszuwerten und zirkulieren zu lassen, erscheinen digitale Informationen als autonom. Diese Autonomie wohnt digitalen Informationen nicht wesentlich inne, sondern ist die Leistung eines Informationssystems, welches die Unabhängigkeit der digitalen Informationen von den sie verarbeitenden Programmen absichert. Dennoch besteht die Gefahr, dass die Bedeutung des Systems vergessen beziehungsweise verdrängt wird und man dem Irrglauben

erliegt, dass es möglich sei digitale Informationen in einer Art kreativen Unordnung halten zu können. Dies propagiert David Weinberger in dem 2008 erschienen Buch *Das Ende der Schublade*. Er vertritt die problematische These, dass die nicht abzustreitende Pluralisierung von Wissensordnungen in digitalen Medien aus der Immaterialität von Bits und Bytes resultiert (vgl. Weinberger 2008, 22). Doch der Pluralität von Präsentations-, Ordnungs- und Handhabungsformen von Informationen an den Benutzeroberflächen des Computers steht nicht die Unordnung von Bits in der Tiefe der Speicher gegenüber. Vielmehr handelt es sich stets um die Verknüpfung und Übersetzung von Ordnungen zwischen Oberfläche und Tiefe. Das World Wide Web, Suchmaschinen, Wikipedia, Soziale Netzwerke, Soziale Taggingssysteme, das Semantic Web etc. stellen Informationsinfrastrukturen bereit, die auf unterschiedlichen Niveaus ansetzen, verschiedenen Logiken folgen und auf unterschiedliche Weise an bestehende Informationsinfrastrukturen anschließen. Die eigentlich bedeutsamen Fragen sind dann, wie im Bereich digitaler Medien unterschiedliche Ordnungen und Verarbeitungslogiken aneinandergesetzt werden, wie sie ineinander übergehen respektive ineinander übersetzt werden und wie diese Übersetzbarkeit sichergestellt wird. Zugleich geraten auch die Bedingungen und Beschränkungen in den Blick, die partikuläre Informationsinfrastrukturen in ihnen gespeicherten Informationen auferlegen sowie die Inkompatibilitäten und Übersetzungsprobleme zwischen unterschiedlichen Formen der Speicherung, Zirkulation und Abfrage von Information.

Anmerkungen

- 01►** Diese Beobachtung schließt an Jacques Derrida an, der auf die Doppeldeutigkeit hinwies, die der Diagnose vom Ende des Buchs innewohnt. Es handle sich um »two extreme, final, eschatic figures of the end of the book, the end as death, or the end as *telos* or achievement« (Derrida 2005, 15). Der antizipierte Tod des Buchs zieht »a constant reinvestment in the book project, in the book of the world, in the absolute book« (Derrida 2005, 15) nach sich, welches sich jedoch in einer anderen medialen Konfiguration zu erfüllen scheint, dem World Wide Web. Kittlers Diagnose bildet hierzu das Spiegelbild. Ausgehend von der Vollendung des Mediums im Computer proklamiert er den Tod der Einzelmedien.
- 02►** Der Rekurs auf Shannons Informationstheorie als zentrale Referenz für eine techniken-zentrierte Medienforschung ist im Anschluss an Kittler zu einem Gemeinplatz medienarchäologischer Forschungsansätze geworden. Besonders deutlich tritt dies in den Arbeiten

von Wolfgang Ernst zu Tage. Kittler und Ernst betrachten Claude Shannon gleichermaßen als den »technical father of modern media culture« (Parikka 2011, 59). Die einseitige Fokussierung auf die nachrichtentechnische Informationstheorie hat zur Folge, dass Medientechniken vorrangig vom Kanal und dessen Indifferenz gegenüber Bedeutung her gedacht werden: »Dabei frappt die Indifferenz technischer Medien, die ebenso diskursive wie nondiskursive, physikalische wie kulturelle Signale buchstäblich gleich-gültig verarbeiten« (Ernst 2008, 163).

- 03►** Die medientheoretische Betrachtung von Computerhardware soll nicht einfach durch die Hinwendung zu Software ersetzt werden. Hardware stellt einen bedeutenden Faktor in der medialen Praxis mit Computern dar. Ihre Bedeutung aber zeigt sich, so die These, erst vor dem Hintergrund konkreter Anwendungen.
- 04►** Zwar mag es stimmen, dass Computer auf der Ebene elementarer Hardware Signale prozessieren, ohne einen Unterschied zwischen Medienobjekten zu machen. Auf der Basis von Hardware werden in der Softwarepraxis, einem Modell geschachtelter Maschinen folgend, jedoch stufenweise immer komplexere Funktionalitäten realisiert, die Unterscheidungen einführen, welche auf einer darunterliegenden Ebene nicht möglich waren. Bedeutsam hieran ist, dass die höheren Programmfunktionen nicht direkt auf die Hardware zurückgeführt werden können, sondern auf die darunterliegende Programmebene: »We conceive an ordered sequence of machines: $A[0]$, $A[1]$, ... $A[n]$, where $A[0]$ is the given hardware machine and where the software of layer i transforms machine $A[i]$ into $A[i+1]$. The software of layer i is defined in terms of machine $A[i]$, it is to be executed by machine $A[i]$, the software of layer i uses machine $A[i]$ to make machine $A[i+1]$ « (Dijkstra 1969, 182).
- 05►** Die Etablierung eines abstrakten und generalisierten Informationsbegriffs bezeichnet Geoffrey C. Bowker als »Information Mythology« (Bowker 1994, 234). Diese bringt die spezifischen Infrastrukturen und Gebrauchskontexte zum Verschwinden bringt, vor deren Hintergrund sich immer erst abzeichnet, was Information ist.
- 06►** Während in den 1980er und 90er Jahren jenseits technikorientierter Ansätze die Immaterialität und Virtualität digitaler Medienobjekte im Vordergrund des Mediendiskurses stand, wird seit Anfang 2000 vermehrt deren fortwährende Materialität und Realität in den Blick genommen.
- 07►** Der Informatiker T. William Olle (Olle 1978, 2), der Linguist Rüdiger Weingarten (Weingarten 1994, 160) und die Kulturwissenschaftlerin Cornelia Vismann (Vismann 2001, 170f.) zitieren gleichermaßen die genannte Definition als historisch ersten Bestimmungsversuch von Datenbanken, ohne auf die Uneinigkeit hinzuweisen, die bezüglich dieses Definitionsvorschlags bei den Teilnehmern der Konferenz bestand.
- 08►** Im Bericht von Franks ist zu lesen: »A broad definition of a data base was brought forward in the following terms leaving unanswered the question of access: A data base consists of: 1. Data[.] 2. Means of access to the data« (Franks 1964, 120).
- 09►** Erhard Schüttelpelz rekonstruiert die Genese des allgemeinen Modells nachrichten-

technischer Kommunikation aus Shannons Hinwendung zum Problem der Geheimplatzkommunikation (Schüttpelz 2002).

- 10 ► Es ist wichtig darauf hinzuweisen, dass Shannons Modell häufig dafür kritisiert wurde, Kommunikation nicht hinreichend zu beschreiben. Diese Kritik richtet sich jedoch eher gegen die Generalisierung des shannonschen Ansatzes als gegen dessen Modellierung des nachrichtentechnischen Kommunikationsproblems. Shannon selbst war sich durchaus der Grenzen seiner eigenen Hinwendung zu Kommunikation und seiner Definition von Information bewusst: »The word ›information‹ has been given many different meanings by various writers in the general field of information theory. It is likely that at least a number of these will prove sufficiently useful in certain applications to deserve further study and permanent recognition. It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field« (Shannon 1953, 105).
- 11 ► In der Informationswissenschaft ist dieses Modell der Informationsspeicherung und Informationsabfrage spätestens seit Ende der 1980er Jahre umstritten, da es das menschliche Suchverhalten nicht adäquat darstellt (vgl. Bates 1989; Morgenroth 2006, 22ff.).
- 12 ► Technologisch wird die Datenbank im Rahmen des Projekts noch nicht als Computerdatenbank entworfen, sondern als eine Mikrofilmbibliothek. Dieser Entwurf weist gewisse Ähnlichkeiten zu Vannevar Bushs Beschreibung der *Memex* sowie zu Emanuel Goldbergs Vision einer Bibliothek der Zukunft auf (vgl. Buckland 1992, 2006). Als früher Vorläufer kann auch das Mundaneum-Projekt von Paul Otlet und Henri La Fontaine angesehen werden. Vgl. hierzu den Aufsatz von Lena Christolova in diesem Band.
- 13 ► Die beiden Formen der Zuschreibung von Bedeutung erläutert Fairthorne am Beispiel des Zählens von Schafen. Um die bereits gezählten Schafe von den übrigen zu unterscheiden, könne man diese entweder markieren oder räumlich separieren: »[W]hen counting sheep, you must distinguish those already accounted for from those not. This can be done with a branding iron [...]. Or you can segregate the counted sheep in a fold« (Fairthorne 1961b [1953], 95).
- 14 ► Die Automatisierung der Suche hat eine zweifache Formatierung zur Folge. Einerseits müssen die versammelten Daten in eine Form gebracht werden, welche es ermöglicht, diese als Information zu adressieren (vgl. Krajewski 2007). Andererseits müssen die Nutzer ihr Informationsbedürfnis gemäß der Adressierungs- und Anfrage-logik der Datenbanktechnologie formulieren, damit die Datenbank korrekte Ergebnisse im Sinn des Informationsbedürfnisses zu Tage fördert (vgl. Weingarten 1988).
- 15 ► Bei Lochkarten zeigt sich die serielle Speicherlogik in deren Anordnung in Stapeln. Gleichwohl eine einzelne Lochkarte ein Datenträger ist, der den Direktzugriff erlaubt, ist der Zugriff auf Lochkarten im Plural seriell organisiert.
- 16 ► Erschwerend tritt hinzu, dass mit der Verfügbarkeit der Festplatte auch alternative Speicherordnungen möglich wurden, was zur Folge hatte, dass die räumliche Anordnung

von Informationen im Speicher zunehmend von ihrer semantischen Ordnung entkoppelt wurde.

- 17► Mit der Markteinführung des Oracle Datenbankmanagementsystems sowie IBMs DB2 Anfang der 1980er Jahre wurde das von Edgar F. Codd entwickelte relationale Datenmodell zum de facto Standard der konzeptuellen Modellierung von Information in Datenbanken (1970).

Bibliografie

- Bachman, Charles W.** (1962a) Concepts of Computer Use in Integrated Systems. In: Charles W. Bachman Papers (CBI 125). Box 20, Integrated Systems Project Notebook #1, Charles Babbage Institute, University of Minnesota, Minneapolis.
- Bachman, Charles W.** (1962b) Integrated Data Store – The Information Processing Machine That We Need! In: Charles W. Bachman Papers (CBI 125). Box 1, Folder 11, Charles Babbage Institute, University of Minnesota, Minneapolis.
- Bates, Marcia J.** (1989) The Design of Browsing and Berrypicking Techniques for the Online Search Interface. In: *Online Review* 13, 5, S. 407-424.
- Bowker, Geoffrey C.** (1994) Information Mythology. The world of / as information. In: *Information acumen. The understanding and use of knowledge in modern business.* Hrsg. v. Lisa Bud-Frierman. London: Routledge, S. 231-247.
- Buckland, Michael** (1992) Emanuel Goldberg, Electronic Document Retrieval, And Vannevar Bush's Memex. In: *Journal of the American Society for Information Science* 43, 4, S. 284-294.
- Buckland, Michael** (2006) Emanuel Goldberg and his knowledge machine. Wesport: Libraries Unlimited.
- CODASYL Data Base Task Group** (1969) Data Base Task Group Report to the CODASYL Programming Language Committee, October 1969. New York: ACM.
- CODASYL Data Base Task Group** (1971) Data Base Task Group Report to the CODASYL programming language committee, April 1971. New York: ACM.
- Codd, Edgar F.** (1970) A Relational Model of Data for Large Shared Data Banks. In: *Communications of the ACM* 13, S. 377-387.
- Derrida, Jacques** (2005) The Book to Come. In: *Paper Machine.* Stanford: Stanford University Press, S. 4-18.
- Dijkstra, Edsger W.** (1969) Complexity controlled by hierarchical ordering of function and variability. In: *Software Engineering. Report of a conference sponsored by the NATO Science Committee, Garmisch, S. 181-185.*

- Ernst, Wolfgang** (2002) Datum und Information. Begriffsverwirrungen. In: Wolfenbütteler Notizen zur Buchgeschichte 27, 2, S. 159-181.
- Ernst, Wolfgang** (2008) ›Merely the Medium?‹. Die operative Verschränktheit von Logik und Materie. In: Was ist ein Medium? Hrsg. v. Stefan Münker & Alexander Roesler. Frankfurt a.M.: Suhrkamp, S. 158-184.
- Fairthorne, Robert Arthur** (1961a) Preface. In: Towards information retrieval. London: Butterworths, S. IX-XI.
- Fairthorne, Robert Arthur** (1961b) Information Theory and Clerical Systems [1953]. In: Towards information retrieval. London: Butterworths, S. 22-41.
- Fairthorne, Robert Arthur** (1961c) The Theory of Communication [1954]. In: Towards information retrieval. London: Butterworths, S. 64-79.
- Fairthorne, Robert Arthur** (1961d) Some Clerical Operations and Languages. In: Towards information retrieval. London: Butterworths, S. 94-105.
- Franks, E. W.** (1964) Session B: Criteria Influencing Data Base Organization or Design. In: Proceedings of the Symposium on Development and Management of a Computer-Centered Data Base (June 10-11, 1963). Hrsg. v. A. Walker. Santa Monica: System Development Corporation, S. 119-121.
- Haigh, Thomas** (2009) How Data Got its Base: Information Storage Software in the 1950s and 1960s. In: IEEE Annals of the History of Computing 31, 4, S. 6-25.
- Hartmann, Frank** (2006) Globale Medienkultur. Technik, Geschichte, Theorien. Wien: WUV.
- Kittler, Friedrich** (1986) Grammophon, Film, Typewriter. Berlin: Brinkmann & Bose.
- Krajewski, Markus** (2007) In Formation. Aufstieg und Fall der Tabelle als Paradigma der Datenverarbeitung. In: Nach Feierabend 2007. Daten; Züricher Jahrbuch für Wissensgeschichte 3. Hrsg. v. David Gugerli, Michael Hagner, Michael Hampe, Barbara Orland, Philipp Sarasin & Jakob Tanner. Berlin: diaphanes, S. 37-55.
- Krämer, Sybille** (2003) ›Schriftbildlichkeit‹ oder: Über eine (fast) vergessene Dimension der Schrift. In: Bild, Schrift, Zahl. Hrsg. v. Sybille Krämer & Horst Bredekamp. München: Fink, S. 157-176.
- Laurel, Brenda** (1993) Computers as theatre. Reading: Addison-Wesley.
- Lesser, Murray L. / Haanstra, J.W.** (1957) The RAMAC Data-Processing Machine. System Organization of the IBM 305. In: Proceedings of the Eastern Joint Computer Conference. New Developments in Computers, New York, S. 139-146.
- Lickliger, Joseph C. R.** (1964) Introduction. In: Proceedings of the Symposium on Development and Management of a Computer-Centered Data Base (June 10-11, 1963). Santa Monica: System Development Corporation, S. 1-7.
- Luhmann, Niklas** (1998) Die Gesellschaft der Gesellschaft. Frankfurt a. M.: Suhrkamp.
- Mooers, Calvin N.** (1950) Information Retrieval Viewed as Temporal Signalling. In: Proceedings of the International Congress of Mathematicians 1, S. 572-573.

- Morgenroth, Karlheinz** (2006) Kontextbasiertes Information Retrieval. Modell, Konzeption und Realisierung kontextbasierter Information Retrieval Systeme. Berlin: Logos Verlag.
- Olle, T. William** (1978) The Codasyl approach to data base management. Chichester: Wiley.
- Parikka, Jussi** (2011) Operative Media Archaeology. Wolfgang Ernst's Materialist Media Diagrammatics. In: *Theory, Culture & Society* 28, 5, S. 52-74.
- Schüttpelz, Erhard** (2002) Eine Ikonographie der Störung. Shannons Flußdiagramm der Kommunikation in ihrem kybernetischen Verlauf. In: *Transkribieren: Medien / Lektüre*. Hrsg. v. Ludwig Jäger & Georg Stanitzek. München: Fink, S. 233-280.
- Shannon, Claude E.** (1948) A Mathematical Theory of Communication. In: *The Bell Systems Technical Journal* 27, 3 & 4, S. 379-423; 623-656.
- Study Group on Data Base Management Systems** (1975) Interim Report. ANSI/X3/SPARC Study Group on Data Base Management Systems. In: *FDT - Bulletin of ACM SIGMOD* 7, 2, S. 1-140.
- Swanson, A. K.** (1964) A Computer-Centered Data Base Serving USAF Personnel Managers. In: *Proceedings of the Symposium on Development and Management of a Computer-Centered Data Base* (June 10-11, 1963). Santa Monica: System Development Corporation.
- Tsichritzis, Dennis / Klug, Anthony** (1978) The ANSI/X3/SPARC DBMS Framework. Report of the Study Group on Database Management Systems. In: *Information Systems* 3, S. 173-191.
- Turing, Alan M.** (1987) Über berechenbare Zahlen mit einer Anwendung auf das Entscheidungsproblem [1937]. In: *Intelligence Service. Schriften*. Berlin: Brinkmann & Bose, S. 17-60.
- Vismann, Cornelia** (2001) Akten. Medientechnik und Recht. Frankfurt a.M.: Fischer.
- von Neumann, John** (1987) First Draft of a Report on the EDVAC [1945]. In: *Papers of John von Neumann on computing and computer theory*. Hrsg. v. William Aspray & Arthur W. Burks. Cambridge: MIT Press, S. 17-82.
- Weinberger, David** (2008) Das Ende der Schublade. Die neue Macht der digitalen Unordnung. München: Hanser.
- Weingarten, Rüdiger** (1988) Das Dialogmuster ›Datenbankanfrage‹. In: *Technisierte Kommunikation*. Hrsg. v. Ders. & Reinhard Fiehler. Opladen: Westdeutscher Verlag, S. 237-243.
- Weingarten, Rüdiger** (1994) Datenbanken. In: *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung*. Hrsg. v. Hartmut Günther & Otto Ludwig. Berlin / New York: Walter de Gruyter, S. 158-170.
- Worsley, Peter K. / Dittman, Roger R. / Weber, Eberhard / Grimberg, Juan C. / Garrett, Peter / Benson-Lehner Corporation** (1959) A Study of the Fundamentals of Information Storage and Retrieval [Final Report]. In: *United States Government Computing Collection, circa 1945-1990 (CBI 63)*. Box 13, Charles Babbage Institute, University of Minnesota, Minneapolis.

»GRAND GAMES OF SOLITAIRE«. TEXTUELLE ORDNUNGEN IN DEN *DIGITAL HUMANITIES*

Ende 2010 stellte Google den *Ngram Viewer* vor, ein experimentelles Projekt aus den unternehmenseigenen *Labs*, mit dem sich Worthäufigkeiten im gesamten Bestand der für Google Books eingescannten Bücher visualisieren lassen. Die Präsentation des Tools fand ein großes Echo in den Medien, nicht zuletzt aufgrund seiner einfachen Bedienbarkeit und der umfangreichen Datenbasis, die angeblich vier Prozent aller jemals gedruckten Bücher abdecken soll. Weniger Beachtung als das Tool selbst fand ein Artikel in der Zeitschrift *Science*, der im Zusammenhang mit der Freischaltung des *Ngram Viewers* für die Öffentlichkeit erschien. In diesem Artikel beschreibt das Entwickler-Team hinter dem Projekt, das aus insgesamt 16 Personen aus verschiedenen Fachbereichen sowie dem Google Books Team besteht, dessen Anwendungsmöglichkeiten in den Kulturwissenschaften (Michel et al. 2011).

Statt in kulturwissenschaftlicher Tradition auf Fragen des Kontexts, der Narration oder der Intertextualität einzugehen, interessieren sich die hauptsächlich naturwissenschaftlich orientierten Entwickler des *Ngram Viewers* allein für die quantitativen Aspekte der analysierten Texte. So zeigen sie zum Beispiel an einem Graphen auf, dass von 1800 bis 2000 der Name ›Freud‹ wesentlich häufiger im allgemeinen Textkorpus zu finden ist als ›Darwin‹, ›Einstein‹ oder ›Galileo‹. Zwar warnen die AutorInnen davor, aus diesen Ergebnissen vorschnelle Schlüsse zu ziehen, dennoch ist die Interpretation der quantitativen *evidence* schnell zur Hand: »›Galileo‹, ›Darwin‹, and ›Einstein‹ may be well-known scientists, but ›Freud‹ is more deeply ingrained in our collective subconscious« (ebd., 182). Auch vor der Anwendung der quantitativen Verfahren auf sehr komplexe historische Themen scheuen die AutorInnen nicht zurück: »For instance, Nazi censorship of the Jewish artist Marc Chagall is evident by comparing the frequency of ›Marc Chagall‹ in English and in German books« (ebd., 181). Der Anspruch des Projekts ist entsprechend hoch gesteckt: Es soll um die Entschlüsselung eines ›kulturellen Codes‹ gehen, der sich – analog zum Mapping von Gensequenzen – quantitativ erfassen lassen soll. In Anlehnung zum Fachgebiet der *genomics*, aus dem ein Teil der EntwicklerInnen stammt, wird daher versucht, den Begriff *culturomics* zu etablieren.

Medien ' Welten

Braunschweiger Schriften zur Medienkultur,

herausgegeben von Rolf F. Nohr

Band 18

Lit Verlag Münster/Hamburg/Berlin/London

Stefan Böhme / Rolf F. Nohr / Serjoscha Wiemer (Hrsg.)

**SORTIEREN, SAMMELN, SUCHEN, SPIELEN.
DIE DATENBANK ALS MEDIALE PRAXIS**

LIT

Bucheinbandgestaltung: Tonia Wiatrowski / Rolf F. Nohr / Stefan Böhme / Serjoscha Wiemer
unter Verwendung diversen Bildmaterials (für Nachweis und © s. Abbildungsverzeichnis),

Buchgestaltung: © Roberta Bergmann, Anne-Luise Janßen, Tonia Wiatrowski

<http://www.tatendrang-design.de>

Satz: Arne Fischer / Rolf F. Nohr

Lektorat: Jasmin Feldberg / Anne Kliche

© Lit Verlag Münster 2012

Grevener Straße / Fresnostraße 2 D-48159 Münster

Tel. 0251-23 50 91 Fax 0251-23 19 72

e-Mail: lit@lit-verlag.de <http://www.lit-verlag.de>

Chausseestr. 128 / 129 D-10115 Berlin

Tel. 030-280 40 880 Fax 030-280 40 882

e-Mail: berlin@lit-verlag.de <http://www.lit-verlag.de/berlin/>

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<http://dnb.ddb.de> abrufbar.

ISBN 978-3-643-11086-2

Printed in Germany

Gedruckt mit Mittel der DFG

(Projekt NO 818/1-1)

DFG

UB LÜNEBURG

0 755 986 0

LÜN 4



Stefan Böhme / Rolf F. Nohr / Serjoscha Wiemer 9
Einleitung

Genealogie & Archäologie der Datenbank

Lena Christolova 31
Das Mundaneum oder das papierne Internet von Paul Otlet
und Henri La Fontaine

Marcus Burkhardt 55
Informationspotentiale. Vom Kommunizieren mit digitalen Datenbanken

Theo Röhle 75
»Grand games of solitaire«. Textuelle Ordnungen
in den Digital Humanities

Uwe Wippich 97
Eugenische Daten – Die Datenpraktiken des Eugenics Record Office

Die Politiken der Datenbank

Martin Warnke 122
Datenbanken als Zitadellen des Web 2.0

Harald Hillgärtner
»Oh, wie süß ist doch die Datenbank!« 137
Zum Aspekt nicht-hegemonialer Datenbanken

Tobias Conradi 159
Prüfen und Bewerten – Redaktionelle Medien
als Gatekeeper der Datenbank?