

Manfred Thaller

**Bemerkungen zu kunsthistorischen
Informationssystemen; vornehmlich aus der Sicht der
Informatik**

Abstract

Der Einsatz von Informationssystemen hat in der Kunstgeschichte sowie auch in anderen Bereichen der Geisteswissenschaften eine lange Tradition. Gesteuert wurde dieser Einsatz in vielen Fällen davon, was zuständige Wissenschaftler als Erfordernisse der Informatik ansahen. Betrachtet man die Informatik etwas genauer, so lässt sich feststellen, dass Prinzipien, die zunächst abstrakt und grundlegend erscheinen, häufig eine ganz bestimmte technische Situation widerspiegeln und ausserhalb eines recht eng definierten Rahmens nicht angewandt werden können. Dies ist vor allem der Fall in Bezug auf Informationssysteme und die Logik der Informationsbeschaffung. Innerhalb dieses Bereiches hat sich der technische Aspekt in den letzten Jahren extrem gewandelt. Es wird deshalb dafür plädiert, dass technische Erfordernisse (welche leicht missverstanden werden können und sich schnell wandeln) nicht als primäre Richtlinien für die Herstellung von kunstgeschichtlichen Informationssystemen dienen sollten, sondern die Erfordernisse für den Einsatz solcher Informationssysteme selbst, die in ihrer Anwendung deutlich werden. Wenn die Informatik als Referenzpunkt für die Herstellung solcher Systeme gelten soll, sollte sie anhand präziser technischer Literatur definiert werden und nicht anhand von Reflexionen über die Bedeutung einiger weniger technischer Begriffe, die leicht missverstanden werden können.

Kontextgebundenheit von Definitionen, Begriffen und Regeln

<1>

In nahezu allen Fächern beginnen die grundlegenden Lehrbücher mit der Definition des Gegenstandes des Faches. Die Informatik macht natürlich im Prinzip keine Ausnahme davon, dennoch fallen diese Definitionen bei ihr meist etwas weniger elaboriert aus, als man bei einem mathematischer Präzision nahe stehenden Fach eigentlich vermuten würde. Dafür gibt es zwei Gründe: Einerseits ist "die Informatik" wesentlich heterogener, als vielfach geglaubt wird. Zwischen Feldern wie der Komplexitätstheorie, die nahezu als Teilgebiet der reinen Mathematik verstanden werden kann und Informatiker hervorbringt, die manchmal ihren Stolz darein setzen, nie in die Nähe eines Computers zu kommen, und der Softwaretechnologie, die sich als reine Ingenieurwissenschaft versteht und eher handwerklichen Stolz vermittelt, bestehen fachintern fast mehr Gegensätze als Gemeinsamkeiten. Dies mag mit der Situation anderer Fächer noch gut vergleichbar sein: fast alle Disziplinen sehen sich selbst wesentlich heterogener, als sie von außen gesehen werden. Andererseits ist die Informatik aber eine der ganz wenigen Disziplinen, die einer Definition ihres Forschungsgegenstandes tendenziell eher ausweichen.

<2>

Dem Außenstehenden liegt die Vermutung nahe, dass unter Kunsthistorikern wohl höchst unterschiedliche Vorstellungen davon bestehen, was 'Kunst' eigentlich sei: Es dürfte aber wohl ein Konsens im Fach darüber bestehen, dass diese Frage wichtig ist. Dies ist in der Informatik, von der man ein ähnliches Verhältnis zum Begriff der Information erwarten könnte, explizit nicht so und es gibt durchaus Lehrbücher, die - in etwa - sagen, Information sei ein schwieriges Konzept, das noch nicht richtig verstanden werde; glücklicherweise wisse man zwar nicht, was sie sei, wohl aber, wie man sie in Rechnern darstellen könne. Und die Lehre davon, wie man sie darstelle und diese Darstellung verarbeite, sei das Fachgebiet der Informatik [1].

<3>

Um ein Missverständnis zu vermeiden: Hier ist die Rede von der Informatik als jener Disziplin, die sich mit tatsächlich existierenden Rechnern und deren Verwendung in einem weiten Anwendungsbereich beschäftigt. In der Nachrichtentechnik [2] gibt es natürlich letzten Endes mechanistische Definitionen von Information die dazu dienen können festzulegen, wie korrekt ein Signal übertragen wird. Man kann allerdings manchmal den Eindruck schwer vermeiden, dass diese Definitionen nicht sehr viel mehr sind als eine sehr präzise Aussage darüber, inwiefern eine Nachricht um so besser übertragen wird, je mehr der gesendeten Bits beim Empfänger unverfälscht ankommen. Dies ist freilich nicht ganz so trivial, wie es anmutet, weil 'übertragene Bits' vielleicht beim Sender und Empfänger nur die Werte Null oder Eins annehmen können, in der dazwischen liegenden Übertragung aber durch Signale weiter gegeben werden, die letztlich sehr viel mehr Zustände annehmen können (dementsprechend sind die Definitionen der Kybernetik zweifellos auch präziser als oben dargestellt, für die praktischen Anwendungen aber dennoch kaum relevant).

<4>

Der merkwürdige Sachverhalt, dass weite Teile einer Disziplin nicht völlig sicher sind, was genau der Stoff ist, mit dem sie sich beschäftigen, sie aber sehr wohl in der Lage sind, anwendbare und oft beeindruckende Ergebnisse zu erzielen, sollte als Warnung verstanden werden: Einerseits davor, das Faszinosum einer fremden Begrifflichkeit als Baugrund für allzu stolze zivilisationstheoretische Gebäude zu verwenden, andererseits um zu betonen, dass 'Regeln', die aus dem zu einem bestimmten Zeitpunkt geltenden State of the Art der Informatik abgeleitet werden um sie für eine andere Disziplin - eben die Kunstgeschichte - nutzbar zumachen, unter Umständen eine erschreckend kurze Halbwertszeit haben können. Daher sollten solche Regeln besser aus einer näheren Betrachtung der Anforderungen der Anwendungsbereiche, konkret der Kunstgeschichte, als aus der Übertragung scheinbar fundamentaler Aussagen 'der Informatik' abgeleitet werden.

<5>

Bestimmte Bereiche der Informatik gehören unstreitig der Mathematik an; insofern sind ihre Aussagen 'zeitlos'. Sobald sie sich jedoch dem Umgang mit der Materie - sprich der Hardware - nähert, sind ihre Aussagen in extremem Maße auf einen kurzlebigen technischen Kontext bezogen. Ein gutes Beispiel dafür ist das Gegensatzpaar 'analog versus digital', das sich sowohl dazu eignet, vor der Errichtung von Theoriegebäuden im begrifflichen Treibsand zu warnen, als auch zu erläutern, warum die unbesehene Übernahme von 'gesicherten' Konzepten zweifelhaft ist.

<6>

Das Begriffspaar 'analog versus digital' kommt ursprünglich aus der Rechnertechnologie [3]. In den Anfangsjahren wurden zwei Techniken verfolgt: Rechner galten als 'analog', wenn sie mathematische Operationen durch physikalische Prozesse abbildeten, die möglichst exakt diese Operation wiedergaben. Werden zwei Spannungen 'addiert', so wird eine Rechenoperation so durchgeführt, dass zwei kontinuierliche Messgrößen - also Messgrößen, die 'beliebig viele Stellen hinter dem Komma' annehmen können - zu einer weiteren kontinuierlichen verbunden werden; wir operieren direkt mit reellen Zahlen. Wird 'digital' gearbeitet, so werden die Operationen auf der Basis von diskreten Messgrößen durchgeführt, also von Messgrößen, die nur eine endliche Anzahl von Ausprägungen - herkömmlicherweise Null und Eins - annehmen können. So weit zur Herkunft der Begriffe in der Rechnertechnologie. Dass 'digital' - durch Ziffern ausgedrückt - mittlerweile als synonym für 'computerbasiert' verwendet wird, ist einfach ein Zeichen für (weit verbreitete) schlampige Begrifflichkeit [4]. Wollte man Zivilisationskritik betreiben, so wäre der 'digitale Sündenfall' irgendwo ins Hochmittelalter [5] zu verlegen. Da sich römische Zahlen zum Darstellen von Kommazahlen ausgesprochen schlecht eignen, waren bis ins Hochmittelalter (und noch lange darüber hinaus) Brüche wie 'zwei Drittel' die korrekte, analoge, konzeptuell infinit exakte Darstellungsweise eines Sachverhaltes, der durch die auf arabischen Zahlen basierende digitale Darstellungsweise '0.666666 ...' selbst durch beliebig viele Kommastellen nur annäherungsweise ausdrückbar ist.

<7>

Freilich ist auch die technische Begrifflichkeit streng genommen 'schlampig'. Was zu einem gegebenen Zeitpunkt im technischen Sinn als 'analog' und 'digital' angenommen wird, ist in wesentlichem Maße eine Frage der zur Verfügung stehenden Messinstrumente. Nehmen wir die Zeit, auf deren Messung das technische Begriffspaar umgangssprachlich vielleicht als erstes übertragen wurde. Eine mechanische Uhr gibt das konzeptuell kontinuierlich teilbare Ablaufen der Zeit durch die vom Zeiger zurückgelegte Strecke wieder, der von einer Unruh bewegt wird, die die in der Feder gespeicherte Energie in kleinen, diskreten Schritten freigibt, also in Wirklichkeit den Zeiger nicht in einer durchgehenden analogen Bewegung weiterschiebt, sondern in voneinander deutlich abgegrenzten Bewegungsschritten, die bei den besten mechanischen Uhren in Bruchteilen von Sekunden angegeben werden. Eine von einem Prozessor betriebene Uhr mit analoger Anzeige entsteht auf digitalem Wege durch die Darstellung einer Zeigerbewegung in kleinen diskreten Schritten [...] gesteuert von einem Taktgeber, dessen Schritte irgendwo zwischen dem Nano- und Millisekundenbereich anzusiedeln sind. Da die Annäherung der rechnergestützten Uhr an einen kontinuierlich aufteilbaren zeitlichen Ablauf somit wesentlich besser ist, als bei der mechanischen Konstruktion, ist die Digitaluhr also erheblich analoger als die mechanische.

<8>

Die Entwicklung der Kunstgeschichte steht, zumindest für den Außenstehenden, in einer engen Wechselbeziehung zur Geschichte der Reproduktion visueller Eindrücke. In jüngster Zeit bestehen diese Wechselbeziehungen nun auch zur sich entwickelnden digitalen Bildverarbeitung, die die Bilder bekanntlich in Pixel, also diskrete, deutlich voneinander geschiedene, Bildpunkte zerteilt, von denen jeder einzelne eine endliche Anzahl deutlich voneinander abgegrenzter Farben annehmen kann. Dies löst eine Phase ab, in der die Kunstgeschichte sich in erster Linie der

Photographie bediente, die ein Bild analog wiedergibt, also als Farbverläufe einer lichtempfindlichen Schicht. Wie aus allen Diskussionen zum Thema 'adäquate Auflösungen für den Einsatz der digitalen Bildverarbeitung' bekannt ist, ist eine der wesentlichen Fragen, die Digitalisierungsprojekte im Bereich der Kunstgeschichte zu lösen haben, welche Körnung die zu digitalisierenden Filme und Fotopapiere haben. Denn eine digitale Auflösung, die feiner ist, als die physikalische Körnung des photographischen Mediums, ist im günstigsten Fall aus Kostengesichtspunkten unzweckmäßig, kann manchmal aber auch die Qualität der Digitalisate beeinträchtigen. Durch die weitere Entwicklung der Bilderfassungs- und -wiedergabegeräte ist damit absehbar, dass die digitalen Bilder kleinere Bildpunkte und glattere Farbverläufe verwenden werden, als die photographischen. Die digitale Bildverarbeitung wird in absehbarer Zeit also der Kunstgeschichte endlich die Möglichkeit wirklich analoger Bildwiedergabe bescheren.

<9>

Wir haben diesen ausführlichen Abschnitt nicht aus Freude an der Polemik vorgeschaltet, sondern um zu zeigen, wie gefährlich es sein kann, aus kontextgebunden präzisen, außerhalb des Kontextes jedoch höchst veränderlichen technischen Termini aus Unverständnis dieser Termini abstrakte Begriffe zu machen, die das 'Wesen' von Techniken beschreiben, oder gar kulturkritische Bedeutung annehmen sollen, für den Informatiker aber bedenklich nach 'wabernder Wortlohe' klingen. Und diese Kontextgebundenheit gilt für sehr viele der informatikbezogenen Begriffe; sie gilt letztlich auch für scheinbar völlig allgemeine Regeln, die zu Zeiten erstellt wurden, zu denen bestimmte technische Rahmenbedingungen so selbstverständlich waren, dass sie nie explizit gemacht wurden.

Information in der Kunstgeschichte

<10>

Bevor wir dies an einem konkreten Beispiel aufzeigen, ein erster Versuch das Thema 'Information in der Kunstgeschichte aus der Sicht der Informatik' einzugrenzen. Ein Hinweis vorab: Ich würde mich dabei gerne dem englischen Sprachgebrauch anschließen, der nur den Begriff der Information, nicht den der Informationen kennt. Begrifflich gesehen treffen wir damit durchaus schon eine Vorentscheidung, die mir geisteswissenschaftlich wichtig zu sein scheint: Informationen im Plural legen nahe, dass es eine letztendlich eindeutig identifizierbare begriffliche Einheit gibt, die vervielfältigbar ist. Wenn Sie so wollen: Informationen legen nahe, dass das dadurch beschriebene Phänomen im obigen Sinne digital ist, also auf wie auch immer geartete Informationseinheiten zurückgeführt werden kann. Information als Begriff, der keinen Plural kennt, beschreibt viel eher ein Kontinuum, innerhalb dessen begriffliche Überschneidungen, Unschärfen und andere Phänomene, wie sie die Geisteswissenschaften kennzeichnen, leichter beschreibbar sind.

<11>

Zunächst ist es sinnvoll, eindeutig zwei Arten von Information zu unterscheiden: Einerseits den Inhalt der Bilder, der als Bitmap gespeichert wird, andererseits die bildbeschreibende Information. Das Bitmap ermöglicht eine Wiedergabe der Farbintensitäten einzelner Bildpunkte auf einer Matrix. Aus Sicht der Informatik ist 'Farbe' ein fast so schwieriges Konzept wie 'Information'. Weder unsere Kenntnis der physiologischen Ausstattung des Homo Sapiens noch unser Verständnis der exakten Prozesse, durch die Farbtöne aus sensitiven Belägen generiert werden, sind

hinreichend präzise, als dass ein bestimmter Farbeindruck wirklich garantiert werden könnte. Mit diesem Vorbehalt können wir aber sagen, dass Bilder das begrifflich einfachere Informationsphänomen im Bereich der Kunstgeschichte sind - sie werden in einer Form gespeichert, die Verläufe von Helligkeits- und Farbphänomenen in materiellen Medien wiedergibt, indem Messgrößen wie 'Helligkeit auf dem roten Farbband' durch eine Zahl repräsentiert werden, die den relativen Abstand dieser spezifischen Helligkeit von den Endpunkten 'größtmögliche' und 'geringstmögliche' Helligkeit wiedergibt (modifiziert durch eine numerische Transformation, die der Tatsache gerecht wird, dass die Unterschiede vom menschlichen Auge nicht linear wahrgenommen werden, sondern physikalisch gleich weit auseinander liegende Helligkeiten in bestimmten Spektral- / Helligkeitsbereichen als unterschiedlich wahrgenommen werden).

<12>

Es fällt nach der Einleitung schwer, nicht darauf hinzuweisen, dass dies eigentlich eine Darstellung per analogiam ist, deren Präzision prinzipiell nur von der Zahl der pro Bildpunkt verwendeten Bits abhängt; wesentlicher ist jedoch, dass es hier um eine Wiedergabe eines in sich homogenen Phänomens geht, das durch einen einheitlichen, keiner intellektuellen Kontrolle im Detail unterliegenden, rein mechanischen Prozess stattfindet.

<13>

Dass diese Art von Information in der Kunstgeschichte sich bisher jenseits der einfachen Reproduktion eines visuellen Eindrucks als relativ sperrig erweist, sei angemerkt [6]: Auch die bekannten Arbeiten von Kirk Martinez in Zusammenarbeit mit Christian Lahanier zur Einbindung von Suchverfahren in eine kunsthistorische Datenbank - im EROS Projekt [7] -, die Bildähnlichkeiten direkt vergleichen, machen nur den Eindruck eines Werkzeugkastens, nicht ausgereifter Verfahren.

<14>

Lassen wir diesen Bereich bei Seite, finden wir uns bei der Bild beschreibenden Information, die wohl über die Jahrzehnte hinweg der Hauptgrund war, warum sich Kunsthistoriker dem Medium Rechner genähert haben: die relativ vage Vorstellung, dieses Medium müsse doch dazu geeignet sein, aus einer großen Menge von kunsthistorischen Belegen den richtigen herauszufinden. Unstrittig ist, dass die dabei betroffene Information im Unterschied zu der eben genannten nicht mechanisch, ohne intellektuelle Intervention, abgebildet werden kann. Der 'Name eines Künstlers' muss explizit bekannt gemacht werden, und selbst wenn er und der 'Name des Kunstwerks' vom Schild auf dem Rahmen des digitalisierten Bildes mechanisch abgelesen werden könnten, würde dies doch nur die technische 'Aktivierung' des intellektuellen Aufwandes bedeuten, der beim Beschriften des Schildes zu treiben war - von der inhaltlichen Beschreibung ganz zu schweigen.

<15>

Bekanntlich gibt es in diesem Bereich einen methodischen Konflikt, der freilich weniger oft offen diskutiert wird, als er dies verdienen würde: Soll sprachliche Information frei formuliert werden oder soll sie einem von mehreren möglichen kontrollierten Vokabularen folgen? Herkömmlicherweise beziehen sich die Befürworter der kontrollierten Vokabulare dabei auf 'die Informatik', genauer gesagt, die Literatur zum Information Retrieval. Fassen wir die übliche Argumentation kurz zusammen.

Information Retrieval: Precision und Recall

<16>

Die Qualität von Information Retrieval Systemen wird herkömmlicherweise durch zwei Maßzahlen bewertet: ‚Precision‘ und ‚Recall‘ [8]. Der ‚Recall‘ bezeichnet den Prozentsatz gefundener Dokumente. Zum Verständnis: Enthält eine Datenbank 100 Beschreibungen von Werken des Künstlers ‚x‘ und eine Abfrage nach Werken dieses Künstlers liefert als Ergebnis 50 davon, so entspricht dies einem Recall von 50 oder 0.5. Die ‚Precision‘ hingegen bezeichnet den Prozentsatz der gefundenen Dokumente, die die Abfrage finden sollte. Erhalten der oder die Suchende im vorigen Beispiel als Ergebnis 100 Dokumente, von denen sich 50 wirklich auf den Künstler ‚x‘ beziehen, so hat diese Abfrage eine Precision von 50 oder 0.5. Der Idealfall wird offensichtlich durch eine Abfrage mit einem Recall von 1 und einer Precision von 1 repräsentiert: Alle Dokumente, die sich auf die Abfrage beziehen werden gefunden, alle gefundenen Dokumente beziehen sich auf das, was gesucht wurde. In beiden Fällen sind Abweichungen unerfreulich: Ein Recall von 0.1 würde andeuten, dass 90 Prozent der vorhandenen, relevanten Dokumente übersehen wurden; eine Precision von 0.1 deutet an, dass 90 Prozent der gefundenen Dokumente irrelevant sind. Dementsprechend geht die klassische Theorie des Information Retrieval davon aus, dass beide Ziele gleichwertig seien, es also genauso wichtig sei, den Recall gegen 1 zu bringen, wie die Precision gegen 1 zu führen.

<17>

Um dies zu erreichen, werden in der Literatur zum Information Retrieval gerne kontrollierte Vokabulare empfohlen, um die bekannten Probleme, die den Recall oder die Precision drücken - Synonyme, orthographische Fehler, Vokabulare unterschiedlicher begrifflicher Trennschärfe - zu vermeiden. In der Regel führt dies zu hierarchischen, sehr fein gegliederten Vokabularen, bei denen implizit die Regel gilt: Was einmal in sehr spezifische Kategorien eingeordnet ist, lässt sich immer maschinell zusammenfassen; aus gröberen Kategorien die feineren automatisch abzuleiten, ist dagegen unmöglich.

<18>

Wie erfolgreich derartige Systeme angewendet werden, ist zweifellos offen für Diskussionen. Ich gebe zu, dass mich etwa angesichts von Iconclass die Tatsache in leichte Zweifel versetzt, dass, um vollständigen Recall bei der Suche nach Fischern zu erzielen, zu beachten ist, dass sie sowohl als

43C121

| | |
|--------|---|
| 4 | Society, Civilization, Culture |
| 43 | recreation, amusement |
| 43C | sports, games and physical performances |
| 43C1 | sports, games ~ animals |
| 43C12 | fishing (sports) |
| 43C121 | angling |

aber auch als

| | |
|-----------|-----------------------------------|
| 47K1(+21) | |
| 4 | Society, Civilization, Culture |
| 47 | crafts and industries |
| 47K | commercial fishery |
| 47K1 | sea and coastal fishery |
| 47K1(+) | KEY |
| 47K1(+2) | labourers ~ crafts and industries |
| 47K1(+21) | men ~ crafts and industries |

beschrieben und gesucht werden können. Ob das wirklich einen soviel besseren Recall produziert, als er zu Stande kommt, wenn man einfach berücksichtigt, dass 'Fischer' auch 'Angler' sein können? Dies sind aber Detailfragen.

<19>

Gravierender ist, dass durch die Festlegung auf möglichst fein gegliederte Beschreibungsschemata einerseits ein sehr hoher Beschreibungsaufwand entsteht - sorgfältige Unterscheidungen zwischen Herkunftsort, Entstehungsort, Fundort, Aufbewahrungsort und eventuell auch noch 'früherem Aufbewahrungsort' in archäologischen Datenbasen -, andererseits aber genau diese diffizilen Beschreibungen voraussetzen, dass dem Suchenden das konzeptuelle Schema vor Augen steht: Soll ein in der Erinnerung vage als 'Pithos aus Samos' abgespeicherter Gegenstand gesucht werden, muss auch wirklich jedes der oben genannten Felder bei der Suche berücksichtigt werden. In den letzten Jahren und Monaten häuft sich nun die Kritik an solchen Systemen als 'benutzerunfreundlich' oder, was ein massiverer Vorwurf ist, als Systeme, die Dinge nicht finden, obwohl sie in ihnen enthalten seien, die also, in der vorhin eingeführten Terminologie, einen sehr schlechten Recall haben.

<20>

Dies ist befremdlich, denn wir sagten ja eben, dass sowohl das Prinzip des kontrollierten Vokabulars als auch das des fein gegliederten Beschreibungsschemas mit Anforderungen der klassischen Untersuchungen zum Information Retrieval begründet wurden, um Recall und Precision zu optimieren. Nun gibt es keine benutzerfreundlichen Systeme, sondern ausschließlich solche, die den Benutzern vertraut sind oder sich wie solche verhalten, die ihnen vertraut sind. Ein wenig salopp gesprochen könnte man die oben ausgesprochenen Beschwerden also so umformulieren: 'Warum finde ich bei Google etwas, wo für die Suche kaum vorbereitete Dokumente behandelt werden, aber nicht in Ihrem System, wo jede einzelne Bildbeschreibung Stunden gekostet hat?'

<21>

Halten wir zunächst fest, dass die beschriebene Situation keineswegs ein Spezifikum kunsthistorischer Datenbanken und Informationssysteme ist. Im Bibliothekswesen enthalten die Datenbanken, die hinter den uns allen vertrauten OPACs stehen, beispielsweise für den 'Titel eines Buches' mindestens folgende Felder (die natürlich auch anders benannt sein können): 'Hauptsachtitel', 'Hauptsachtitel in Ansetzungsform', 'Hauptsachtitel Zusatz', 'Nebeneintrag', 'Nebeneintrag in Ansetzungsform' und 'Gesamttitle', die einen kleinen Teile der sechzig oder hundert Felder ausmachen, mit denen jedes Buch beschrieben wird. Das mag geübte OPAC-

BenutzerInnen wundern, weil sie natürlich gewohnt sind, einfach nach einem 'Titel' zu fragen. Das Rätsel löst sich dadurch, dass die erste Aufgabe jedes Softwaredesigners, der einen OPAC oder ein ähnliches System konstruiert, darin besteht, die zahllosen Kategorien, in die die Kataloginformation der Bibliotheken in den Datenbanken aufgeteilt wird - Recall and Precision! - zu wesentlich weniger präzise formulierten Kategorien zusammenzufassen, die sich dem bibliographisch weniger präzisen Denken der OPAC-BenutzerInnen anpassen.

<22>

Nun ist es schwierig, angesichts dieser Situation nicht die Frage aufzuwerfen, ob die Aufspaltung der Information auf diese vielen Felder denn wirklich unbedingt notwendig sei, wenn sie, sobald sie dem menschlichen Benutzer präsentiert werden, wieder zu vageren Kategorien zusammenfließen. In der Tat gibt es in der Welt der Bibliothekare mittlerweile einen Trend zu wesentlich weniger trennscharfen Systemen. So baut der Dublin Core [9] zunächst auf der Vorstellung auf, dass alle Bücher dieser Welt - und streng genommen alle Objekte, für die Bibliotheken verantwortlich sind, durch fünfzehn Felder beschrieben werden können: 'Title', 'Creator', 'Subject', 'Date' und andere, sehr allgemeine Begriffe. Die Verbindung zwischen den elaborierten früheren Schemata wird dadurch hergestellt, dass das Grundkonzept 'Titel' ergänzt werden kann durch einen Zusatz, der angibt, dass es sich bei diesem 'Titel' um das funktionale Äquivalent eines 'Haupttitels' oder eines 'Nebeneintrags' handelt. Noch wesentlich radikalere Vorschläge existieren: So sieht das System der Electronic Resource Citation [10] noch genau vier Felder vor - 'who', 'what', 'when' und 'where' die angeblich für eine 'Minimalbeschreibung jedes Objekts im Universum' reichen (und gegebenenfalls durch eine stufenweise hierarchische Verfeinerung ihrerseits wieder angepasst werden können).

<23>

Dieser Ausflug in die Welt der Bibliotheken schien angebracht, weil die heutigen Informationssysteme der Bibliotheken - im Unterschied zu denen der Kunstgeschichte - in der Entwicklung des Information Retrieval selbst eine sehr zentrale Rolle gespielt haben, also ursächlich beteiligt waren an der Entstehung des folgenden Lehrsatzes: 'um Recall und Precision gleichzeitig zu optimieren, sind möglichst fein gegliederte Beschreibungsschemata mit möglichst fein gegliederten kontrollierten Vokabularen nötig.'

<24>

Nun haben wir aber einleitend am Beispiel des Begriffspaars 'analog versus digital' gezeigt und davor gewarnt, dass scheinbar abstrakte Konzepte der Informatik in Wirklichkeit nur innerhalb eines recht begrenzten Kontexts aktueller technischer Entwicklung sinnvoll sind. In welchem technischen Kontext ist die 'Lehre von Recall and Precision' entstanden? Die Informationssysteme, in deren Kontext die bestehenden Konzepte des Information Retrieval entwickelt wurden, waren meist Systeme der allerersten Generation von Online-Systemen. Das bedeutet, dass (a) nach Formulierung einer Abfrage eine zum Teil nicht unerhebliche Wartezeit - eventuell bemessen in Stunden - eintrat, (b) entweder das gesamte Suchergebnis in einem Stück auf Papier ausgedruckt werden musste oder mit einer weiteren Wartezeit für den Rest eines nur angedruckten Ergebnisses zu rechnen war und (c) sowohl die Verbindungszeit mit dem Informationssystem als auch der Umfang der Ergebnisse erheblich als Kostenfaktoren zu Buche schlugen.

<25>

Dies führte zu dem Bestreben, mit der ersten Anfrage möglichst genau jenes Ergebnis zu erzielen, das tatsächlich benötigt wurde. Ein nicht benötigtes Ergebnis mit einigen tausend Treffern hätte nicht nur Papier verschwendet, sondern durch lange Verbindungszeiten mit dem Informationssystem auch direkt hohe Kosten verursacht. Deshalb waren die meisten Informationssysteme dieser ersten technischen Generationen von Anfang an nie für den 'Laien' konzipiert. Das ganze Berufsbild des Dokumentars / der Dokumentarin entstand in nicht unerheblichem Maße aus der Überzeugung, dass der kosteneffektive Umgang mit elektronischen Informationssystemen so schwierig sei, dass man zwischen BenutzerIn und Informationssystem einen Dokumentar oder eine Dokumentarin schalten sollte. Diese müsse in mindestens ein- bis zweijähriger Ausbildung das Formulieren effizienter Abfragen genauso lernen wie die 'Pfleger' der erforderlichen stark gegliederten und terminologisch sorgfältig durchgearbeiteten Datenbanken, mit anderen Worten die Dateneingabe. In genau dieser technischen Situation entstanden die Strategien, die zur gleichzeitigen Optimierung von 'Recall' und 'Precision' führten, die implizit noch heute bei der Anlage vieler kunsthistorischer Informationssysteme mitwirken.

<26>

Nichts an diesem Szenario entspricht heute noch der Realität: Die Wartezeiten von WWW-basierten Informationssystemen werden in Sekunden bemessen, Ergebnisse werden in Segmenten ausgegeben und das zunächst auf dem Bildschirm, also auf einem Medium, bei dem es völlig gleichgültig ist, ob Hunderte irrelevanter Angaben auftauchen. Auch die Verbindungszeiten sind heute kein wirklich kritischer Kostenfaktor mehr. Da die Gesamtkosten des Zugangs zu den Informationssystemen sinken, wird es aber gleichzeitig nahezu unvorstellbar, eine ausgebildete Fachkraft mit der Formulierung einzelner Informationssuchen zu beschäftigen. In der einschlägigen Fachliteratur führt dies dazu, dass zwar die Optimierung von Recall und Precision noch als einführendes Thema gelehrt wird - natürlich sind viele der dazu entwickelten Verfahren und Techniken weiter nützlich - , darauf aufbauend wird aber recht nachdrücklich darauf hingewiesen, dass es derzeit eigentlich noch keine wirklich klaren Methoden gäbe, um die Qualität der Reaktion von Informationssystemen wie Google oder anderen WWW-Suchmaschinen zu bewerten [11].

Arbeitsteilung zwischen Mensch und Maschine

<27>

In der neueren Literatur zum Thema Information Retrieval zeichnet sich also ein Trend ab - auch wenn dies in exakt diesen Worten bisher kaum ausgesprochen wird: die Trennung der Arbeit zwischen Mensch und Maschine. Der Rechner als solcher ist ungemein geeignet, um Routineaufgaben mit nicht ermüdender Aufmerksamkeit durchzuführen. Dementsprechend bekommt er die Aufgabe, auf Grund vergleichsweise unpräziser Anfragen wirklich alles gesuchte Bildmaterial anzuzeigen. 'Gesucht sind alle Bilder von Rembrandt' heißt dann nicht 'suche alle Dokumente, bei denen im Feld 'Künstler' die Zeichenkette 'Rembrandt van Rijn' enthalten ist', sondern etwas Ähnliches wie 'suche alle Dokument, bei denen ein beliebiges Feld den Text 'Rembrandt' enthält - und gib Dich unter Umständen auch zufrieden, wenn dort 'Rembrand' steht, falls es eine europäische Sprache geben sollte, in der dies die geläufigere Notation ist'. Anders ausgedrückt: Der Rechner optimiert den Recall - selbstverständlich ist die gefundene Liste von Belegen alles andere als 'korrekt'.

Mindestens enthält sie die Werke der Schüler; je nach verwendeten Datenbanken sogar Werke, die nach Ansicht des Beschreibenden Ähnlichkeiten mit dem Werk des Genannten haben. Aus diesem Material aber, angeboten als Liste, wählt der menschliche Benutzer die relevanten Dokumente aus: Denn genau hier, in der raschen Sichtung grob vorsortierten Materials nach Kriterien, die dem Bearbeiter oder der Bearbeiterin meist selbst nicht so richtig klar sind, bevor sie sich in der Auseinandersetzung mit dem verfügbaren Corpus präzisieren, liegt die Stärke des menschlichen Bearbeiters. Er (oder sie) ist dann für die Precision verantwortlich [12].

<28>

Diejenigen Leser, die seit längerer Zeit mit dem Sujet befasst sind, werden sich noch eines weiteren, ständig wiederkehrenden Themas bei den Diskussionen über geisteswissenschaftliche Datenbanken in den Siebzigern, Achtzigern und frühen Neunzigern erinnern, das, immer nur als Argument von Skeptikern in die mündlichen Diskussionen nach Vorträgen eingebracht, kaum je den Weg in den Druck fand: Die Ergebnisse einer datenbankgestützten Suche beispielsweise in einem der Vorläufersysteme der heutigen OPACs (und noch mehr in allen Datenbanken, die inhaltlich gewichtigeren Katalogen zu entsprechen versprochen) seien ja sehr beachtlich, ging das Argument, aber selbst wenn alle Kinderkrankheiten ausgemerzt wären - der jeweilige Debattenredner selbst würde doch das Erlebnis beim Konsultieren gedruckter Nachschlagewerke vermissen, das einem den Zugriff auf Information bringe, die man nicht gezielt gesucht hätte, weil man sie nicht kannte, auf die man auf Grund der Organisation der Nachschlagewerke aber stoße, weil sie einem beim Blättern als dem eigentlich Gesuchten benachbart auffiele.

Vom Information Retrieval zum Information Browsing

<29>

Die Lösung scheint in einem Wechsel vom Primat des Information Retrieval zum Information Browsing zu liegen, als der den begriffliche 'weichen' Geisteswissenschaften angemessensten Zugangsform. Damit würde der bisher gerne als sozusagen 'archetypisch' geltende Zugriff mittels sechs mit booleschen Operatoren verbundener Prädikate jenen Fällen vorbehalten bleiben, die der gezielten Nachsuche nach einem nicht präzise genug zitierten Einzelstück gelten, während der neue Archetypus nicht so sehr das buchhalterische Blättern im Zettelkasten, als das eigentlich kreative Querlesen in der Bibliothek implementiert. Die Konsequenzen wären nicht unerheblich: So würde etwa hinsichtlich der peripheren, aber hoch angesehenen Denkrichtung um Jean Claude Gardin, die den primären Nutzen der Rechnertechnologie für die Geisteswissenschaften nicht in der Produktion von Ergebnissen, sondern in der Entwicklung geisteswissenschaftlicher Begrifflichkeiten sieht, die hinreichend präzise wären, um mit dem Rechner bearbeitbar zu sein [13], klar, warum, bei aller Faszination, die diese Arbeiten häufig ausstrahlen, so erstaunlich wenige Forscherinnen und Forscher bereit waren, sich dieser Aufgabenstellung ernsthaft zu widmen. Was die derzeitige Entwicklung der theoretischen Beschäftigung mit Informationssystemen betrifft, ist jedenfalls festzuhalten, dass sie diesem Schritt vom 'Suchen' zum 'Blättern' keinesfalls ablehnend gegenübersteht [14].

<30>

Freilich, wenn dies so ist, wenn die Suche des Rechners die verstehende menschliche Sichtung nur vorbereiten und optimieren, auch konzeptuell nicht

ersetzen soll, ist dann selbst unsere - konventionelle - einleitende Trennung kunsthistorischer Information in die 'nicht standardisierte' bildliche Information und die in präzisen Beschreibungen niedergelegte sprachliche Information noch ganz so relevant? Denn: der menschliche Partner in unserem Modell von einer 'Mensch-Maschine-Kooperation zur Auswahl des benötigten Materials' profitiert ja unstreitig sehr von der Präsenz der bildlichen Information um zu entscheiden, welche Belege er oder sie für das Ziel, das den aktuellen Suchvorgang motivierte, eigentlich braucht. Dies wird deutlicher, wenn wir dieses Ziel nochmals näher betrachten.

<31>

Wir haben oben schon gesagt, dass die klassischen Lehren vom Aufbau von Information Retrieval-Systemen von bestimmten selbstverständlichen Voraussetzungen ausgingen, die zum Teil durch die geänderten technischen Gegebenheiten obsolet geworden seien. Wenn wir die letzten Absätze ernst nehmen, könnten noch sehr viel mehr dieser Grundannahmen für Informationssysteme für die Geisteswissenschaften im Allgemeinen und die Kunstgeschichte im Besonderen obsolet sein. Wenn wir davon ausgehen, dass die bildliche Information für den menschlichen Benutzer ein ganz wesentliches Kriterium für die Entscheidung sei, ob der gefundene Beleg - im Falle kunsthistorischer Datenbanken also die Abbildung - benötigt werde, sollten wir nochmals der Frage 'benötigt wofür' nachgehen.

<32>

Im Falle der meisten natur- oder ingenieurwissenschaftlichen Datenbanken ist der Sinn einer Datenbankabfrage völlig klar. Wenn eine bestimmte chemische Verbindung für einen bestimmten Einsatzzweck zu kostspielig, da zu schwierig herzustellen ist, sind wir an Verbindungen mit möglichst ähnlichen materiellen Eigenschaften interessiert, die kostengünstiger sind. Dementsprechend wäre es undenkbar, wenn wir im ingenieurwissenschaftlichen Fall nach chemischen Verbindungen suchen, die eine Oberfläche hitzeresistenter machen und dabei auf Material stoßen, das die Spröde bei Kälte reduziert. Dem entspricht natürlich auch die einschlägige Fachliteratur: Ingenieurwissenschaftliche Lehrbücher handeln die Eigenschaften bestimmter Materialgruppen ab, ohne Assoziationen zu andersartigen Materialien herzustellen - was letzten Endes der Grund ist, warum natur- und ingenieurwissenschaftliche Publikationen aus geisteswissenschaftlicher Sicht so ungemein knapp und spröde sind: weil sie jeweils genau einen abgeschlossenen Sachverhalt beschreiben.

<33>

Geisteswissenschaftliche Publikationen mögen manchmal schlecht geschrieben sein, aber knapp und spröde sollten sie eigentlich nie werden, denn auch für den Autor, der sich über die Entwicklung des modernen Designs auslässt, ist die Beobachtung, dass ein vage ähnlicher Effekt sich bei der Entwicklung minoischer Wandmalereien eingestellt habe, keineswegs notwendigerweise abseitig: Ja, die Herstellung nicht vermuteter Verbindungen ist oft ein so wünschbarer Effekt geisteswissenschaftlicher Forschung, dass solche Verbindungen allzu oft mühsam herbei argumentiert werden und dann gequält wirken. Wenn die Art und Weise, wie begriffliche Nähe und Ferne in den geistes- und nicht-geisteswissenschaftlichen Fächern in der Literatur und in der Lehre gehandhabt wird, sich so stark unterscheidet, warum müssen dann Informationssysteme aus diesen beiden Welten versuchen, sich dieselben Strukturprinzipien zu eigen zu machen? Weil die Wissenschaft von diesen Systemen

sie vorschreibt?

<34>

Sie schreibt es deshalb vor, weil sie zwei Voraussetzungen macht bzw. machte: (a) Es existiert eine völlig eindeutige 'richtige' Antwort für die Suche und (b) die Suche führt zu Schlüsselinformation, die dazu geeignet ist, Zugang zum gesuchten Objekt zu bekommen, es aber nicht ersetzen kann. Dass (a) in den Geisteswissenschaften in dieser Form nicht zutrifft, haben wir oben darzulegen versucht; auch im Falle von (b) wird diese Voraussetzung auch und gerade in den hochauflösenden, also qualitativ hochstehenden, Bilddatenbanken der Kunstgeschichte / der Geisteswissenschaften rasch obsolet. Einem Ingenieur, dem der Zugriff auf die Basisinformation über eine chemische Verbindung angeboten wird, erspart diese Information den Zugriff auf die Verbindung selbst - um sie in den geplanten experimentellen oder Produktionsprozess einzubauen - keineswegs. Und natürlich ersetzt auch eine vorzügliche Bildschirmabbildung des Petersdoms das psychische Erlebnis des ersten Moments in dem die Dimensionen des Raumes wirklich klar werden, keinesfalls. Aber: auch das Nachschlagen in einem Essay, in dem der Petersdom an Hand umfangreichen Abbildungsmaterials diskutiert wird, ersetzt den Besuch nicht.

<35>

Das Bild auf dem Bildschirm kann aber sehr wohl den Zugriff auf den Bildband überflüssig machen - wenn die jenseits der eigentlichen Abbildung im Essay enthaltene Information zusammen mit den Bildern dargestellt werden kann. Und: Zu beurteilen, ob die kunsthistorische Fachdebatte in erster Linie eine Debatte über Kunstwerke (den Petersdom) oder eine über kunsthistorische Interpretationen (die Literatur dazu) sei, steht dem Nichtkunsthistoriker nicht zu; dem benachbarten Auch-Geisteswissenschaftler sei ein gewisser Verdacht aber nachgesehen.

<36>

Wir gehen daher davon aus, dass, wie durch die bisherigen Überlegungen naheliegend, Geisteswissenschaftler im Allgemeinen und Kunsthistoriker im Besonderen durch Bilddatenbanken in die Lage versetzt werden können, durch den Zugriff auf ein Bild, umgeben von einer essayistischen Beschreibung, auf den Zugriff auf das klassische Medium des Bildbandes insgesamt zu verzichten. Während klassische Datenbanken aus den aufgezeigten Gründen also versucht haben, Beschreibungen zu generieren, die als Suchwerkzeuge für den Rechner dienen, um Benutzer und Benutzerin davon zu informieren, welche Darstellungen sie in traditionelleren Medien nachschlagen sollen, wäre eine Verbindung von qualitativ hochstehenden Reproduktionen mit qualitativ hochwertigen Beschreibungen, die nicht so sehr ein Findmittel zum begleitenden Bild, sondern eine Auseinandersetzung mit diesem Bild darstellen, also eine Möglichkeit, das Ergebnis der Suche in einem Informationssystem unmittelbar nutzen zu können; um in einen komplexen Kontext einzutauchen, der Beziehungen zwischen den behandelten Bildern wiedergibt.

<37>

Das aber bedeutet, dass wir uns von der Vorstellung einer 'Datenbank' als einer Sammlung von (ursprünglich sekundären oder überhaupt nicht präsenten) Bildern, die durch eine streng durchgeplante Anzahl von Feldern mit streng kontrolliertem Vokabular beschrieben werden, verabschieden würden. Ersetzt würde eine derartige Datenbank durch Informationssysteme, die eine verblüffende Ähnlichkeit mit den

klassischen Ausstellungskatalogen haben könnten: knappe und relativ präzise formulierte Bildunterschriften im Kontext vergleichsweise umfangreicher, frei formulierter Texte [15]. Bearbeitbar einerseits durch Suchmaschinen, die in den frei formulierten Texten suchen (wobei Thesauri im Hintergrund relativ leicht Synonymverzeichnisse verwalten könnten); bearbeitbar aber auch durch Teile dieser Suchmaschinen, die gezielt auf jene Teile des Textes zugreifen, bei denen mit einer größeren terminologischen Präzision zu rechnen ist - zum Beispiel dem Äquivalent der knappen Bildunterschriften, die ja eine enge Beziehung zu den oben geschilderten 'schlanken' Kategoriensystemen aus dem bibliothekarischen Bereich haben.

<38>

Lassen Sie mich die obige Argumentationskette nochmals zusammenfassen, bevor ich zu einer abschließenden Beobachtung komme.

Kunsthistorische Informationssysteme, wie wir sie heute kennen, spiegeln einen ganz bestimmten Entwicklungszustand der Informatik - oder genauer gesagt jenes Teiles von ihr, der sich mit dem Information Retrieval beschäftigt - wider. Da in dieser Phase der technischen Entwicklung die Beschäftigung mit sehr regelmäßigen Daten und sehr starker konzeptueller Kontrolle unterliegender Information im Vordergrund stand, entwickelte man das Konzept einer Datenbank, die sehr präzise Suchen nach genau bekannten Objekten unterstützen sollte, die nahezu vollkommen unabhängig voneinander waren.

<39>

Alle Konzepte der Informatik, die auch nur im weitesten Sinne mit dem Stand der technischen Entwicklung verbunden sind, sind nur innerhalb dieses Standes der technischen Entwicklung gültig. Wir haben einleitend mit voller Absicht gezeigt, dass dies auch - und gerade! - für so grundlegende Konzepte wie 'digital versus analog' gilt. Wird eine Technik, die aus einem bestimmten Entwicklungsstand der Informatik herrührt, nachdem dieser Stand bereits überholt ist, weiterhin zu Grunde gelegt, so ist dies nur sinnvoll, wenn das auf diesem Entwicklungsstand favorisierte 'Geschäftsmodell' innerhalb einer bestimmten Wissensdomäne aus ihr immanenten Gründen weiterhin wertvoll ist.

<40>

Ob dies für das Geschäftsmodell 'präzise Suche nach knapper, kontextfreier Information' in der Kunstgeschichte zutrifft, scheint mir sehr zweifelhaft. Der geisteswissenschaftlichen Tradition scheint es wesentlich besser zu entsprechen, sich geisteswissenschaftliche Informationssysteme als Netze lose zusammenhängender Informationsblöcke vorzustellen, innerhalb derer einzelne Punkte mit Hilfe der Informationstechnik rasch gefunden werden können und dann so dargestellt werden, dass der Benutzer oder die Benutzerin rasch entscheiden kann, ob dieser Punkt des Informationsnetzes die tiefer gehende Beschäftigung, über das 'Durchblättern' der umgebenden Seiten, rechtfertigt.

<41>

Die eben angestellte Vermutung, welche Strukturen für kunsthistorische Informationssysteme besonders angemessen seien, hätte schon früher aufgestellt werden können - und wurde dies auch. Sie gewinnt aber heute besonderes Gewicht, weil genau diese Art von Struktur, vor wenigen Jahren noch sehr exotisch, mindestens vorübergehend ins Zentrum der Beschäftigung mit Informationssystemen

gerückt ist. Die völlig neue Bedeutung, die 'Interaktivität' in den letzten Jahren angenommen hat, lässt die Bedeutung sehr grundlegender Messgrößen der Theorie des Information Retrieval völlig neu bewerten. Wir könnten darauf hinweisen, dass sich im Umfeld der aus der Beschäftigung mit den netzartigen Strukturen des WWW resultierenden Technologien eine ganze Reihe von Techniken herausgebildet hat bzw. herausbildet, die gerade das, was vor wenigen Jahren noch exotisch gewesen wäre, technisch gut handhabbar macht.

<42>

Dementsprechend nochmals und explizit ein abschließender Appell: Die Entwicklung der Informatik in ihrer angewandten Form, mindestens die Entwicklung der Informationstechnologie in den letzten fünf Jahre, hat zu Techniken und Konzepten geführt, die wesentlich näher an plausiblen Szenarien für den Einsatz der Informationstechnik in der Kunstgeschichte liegen als frühere [16]. Dies wäre also ein sehr guter Zeitpunkt, um sich nochmals verstärkt der Frage zuzuwenden, was die Informationstechnologie für die Kunstgeschichte im Idealfall bewirken sollte [17], ohne sich durch allzu viele vorgefasste Meinungen darüber behindern zu lassen, was diese Technologien angeblich für Voraussetzungen haben [18]: Diese 'Voraussetzungen' stellen sich allzu oft als terminologischer Treibsand nach Art des 'fundamentalen Unterschiedes' zwischen 'digital' und 'analog' heraus, mit dem wir uns einleitend beschäftigt haben.

Anmerkungen

- 1 Ulrich Rembold / Paul Levi: Einführung in die Informatik für Naturwissenschaftler und Ingenieure, München / Wien 1999, 41, erklärt Information zu einem nicht weiter definierten, sondern nur durch Eigenschaften beschriebenen Grundbegriff. Dazu zähle beispielsweise die Notwendigkeit, sie zu kodieren. Natürlich ist dies ein extremes Beispiel und Texte wie Manfred Broy: Informatik, I, Berlin 1998, 3, betonen die Zentralität des Verständnisses von Information für die Informatik sehr wohl; auch hier wird aber nachhaltig darauf hingewiesen, dass sich die Informatik nur der Frage der Repräsentation der Information, nicht der Information selbst annehme und annehmen könne.
- 2 Claude E. Shannon hat in seinem grundlegenden Aufsatz: A Mathematical Theory of Communication, in: Bell System Technical Journal 27 (1948), 379-423, 623-656, den gerne zitierten Satz geprägt: "whatever resolves uncertainty is information". Als Ausgangspunkt für eine damals erst beginnende Diskussion unstreitig eine brillante Formulierung. Die unmittelbare technische Relevanz des Zitats wird aber vielleicht am besten dadurch beleuchtet, dass Google bei einer Suche nach diesem Zitat ungefähr genau so viele Vorlesungsskripten aus Bereichen der Philosophie, Anthropologie, Theologie etc. findet, die dieses Zitat kommentieren, wie solche aus der Informatik / Informationsverarbeitung.
- 3 Vergleiche Rembold / Levi: Einführung, 42, zur Unabhängigkeit von "diskreter" versus "kontinuierlicher" Information und ihrer Darstellung.
- 4 Ein klassisches Beispiel für die dabei auftretenden terminologischen Wirrnisse ist der Aufsatz von Andre Reifenrath: Relation und Realität. Von den Problemen der Informationsabbildung in elektronischen Systemen, in: Hubertus Kohle (Hg.): Kunstgeschichte digital, Berlin 1997, 27-41. Siehe dazu auch die letzte Anmerkung zu diesem Aufsatz.

- 5 In diesem Kontext muss natürlich auf das Gesamtwerk Harald Witthöfts verwiesen werden, der mit seiner Herausarbeitung einer "historischen Metrologie" unter anderem die Forschungsthese vertritt, dass das antike-mittelalterliche Zahlendenken in präzisen Proportionen organisiert war, demgegenüber unser "Zahlendenken" zunächst einen konzeptuellen Verlust an Präzision ausmacht. Wie immer bei Aussagen, die sich auf ein Lebenswerk beziehen, ist ein passendes Zitat schwer auszuwählen. Für den Interessenten sei als Ausgangspunkt daher sein frühes Hauptwerk zitiert: Harald Witthöft: Umriss einer historischen Metrologie zum Nutzen der wirtschafts- und sozialgeschichtlichen Forschung (= Veröffentlichungen des Max-Planck-Instituts für Geschichte 68), Göttingen 1969, Band I und II.
- 6 Eine heilsam ernüchternde Einschätzung der Möglichkeiten erhält man zum Beispiel wenn man die skeptische Bewertung bei Ian H. Witten u. a.: Managing Gigabytes, New York 1999, 444, gerade im Kontext der dort diskutierten hochkomplexen Verfahren in anderen Bereichen der Informationssysteme betrachtet.
- 7 Leider noch nicht in gut zitierbarer Literatur dokumentiert. Zum Zeitpunkt der Veröffentlichung als erster Einstieg: Christian Lahanier: Eros. Open Source Multilingual Research System, http://www.mmi.unimaas.nl/eculturenet/publicPDF/lahanier_EROS.pdf (02.05.2003).
- 8 Für die klassische Herleitung vergleiche Ricardo Baeza-Yates / Berthier Ribeiro-Neto: Modern Information retrieval, Boston u. a. 1999, 73. Dabei sollte beachtet werden, dass auch außerhalb der weiter unten zitierten Kontexte diese Maßzahlen nicht annähernd so problemlos sind, wie sie in populären Übernahmen für geisteswissenschaftliche Auditorien gerne zitiert werden: Baeza-Yates / Ribeiro-Neto, Modern Information retrieval, 81.
- 9 Wie alle internationalen Standardisierungsunternehmen dieser Art am besten dokumentiert durch die Website: <http://www.dublincore.org/> (02.05.2003).
- 10 DC-Kernel / ERC Special Interest Group, <http://dublincore.org/groups/kernel/> (02.05.2003)
- 11 Vergleiche Baeza-Yates / Ribeiro-Neto: Modern Information retrieval, 261, zur Problematik der Bewertung interaktiver Systeme durch die klassischen Maße. Allgemein zur Problematik der WWW-Szenarios für die klassischen IR-Ansätze, ebenda 367.
- 12 Vergleiche Baeza-Yates / Ribeiro-Neto: Modern Information retrieval, 117, zu einer anderen Strategie, bei der die allgemein bekannte Problematik den ersten Zugang zu finden, dadurch umgangen wird, dass der Rechner die Informationssuche insgesamt als stufenweisen Prozess unterstützt und versucht aus Benutzerreaktionen zu lernen.
- 13 Meine Paraphrase von Jean Claude Gardin: Le Calcul et La Raison, Paris 1991.
- 14 Vergleiche Baeza-Yates / Ribeiro-Neto: Modern Information retrieval, 65. Zum "Browsen" als einem eben so voll in eine Gesamtheorie des Information Retrieval integrierte Technik wie die klassische Suche. Dort auch Seite 394 zur Verbindung beider Strategien als einer zentralen künftigen Entwicklungsaufgabe.
- 15 Wobei hier generell auf die sich entwickelnde Technologie des Data Mining zu verweisen ist. Dazu - und zum Verhältnis dieser Techniken zu den IR-Techniken, die die kunsthistorischen Informationssysteme implizit beherrschen, - zum Beispiel Jiawei Han / Micheline Kamber: Data Mining, San

- Francisco u. a. 2001, in unserem Kontext insbesondere Seite 428.
- 16 Die Bezüge des Folgenden mögen für viele Leser dieser Seiten nicht offensichtlich sein. Nichtsdestoweniger sei auf Kevin Williams u. a.: Professional XML Databases, Birmingham 2000, 409, hingewiesen. Was dort vorgeschlagen wird, läuft letzten Endes darauf hinaus, dass jedes XML Dokument - und damit letzten Endes jeder Text überhaupt - potentiell von einem Datenbanksystem verwaltet werden kann.
- 17 Übrigens ist es eine fundamentale Forderung lehrbuchmäßigen Softwaredesigns zunächst ein Modell des Fachkonzepts zu erstellen und dabei alle Aspekte der Implementierung auszuklammern. Jedes Mal, wenn ein Kunsthistoriker oder seine Kollegin sich ganz besonders an den "bekanntesten technischen" Gegebenheiten orientieren, begehen Sie einen Bruch der fundamentalsten softwaretechnischen Voraussetzungen: Vergleiche Heide Balzert, Lehrbuch der Objektmodellierung, Heidelberg / Berlin 1999, 8.
- 18 Ein klassisches Beispiel in dieser Hinsicht ist, wie gesagt, Andre Reifenrath, Relation und Realität. Von den Problemen der Informationsabbildung in elektronischen Systemen, in: Hubertus Kohle (Hg.): Kunstgeschichte digital, Berlin 1997, 27-41. Hier wird, durch schlichte und fundamentale Unkenntnis dessen, was die Informatik jenseits der Anwenderhandbücher von Systemen wie Microsoft Access anbietet, ein Popanz aufgebaut, der ebenso unrealistisch ist, wie er bekämpfenswert wäre, wenn er der Realität entspräche. Vergleiche Seite 34: "Entsprechend dem Kalkül ihrer Boole'schen Operationen kennen sie nur vorhandene Objekte oder nicht vorhandene Objekte." In Information Retrieval-Lehrbüchern, wie dem hier häufig zitierten von Baeza-Yates und Ribeiro-Neto werden etwa 20 unterschiedliche Vorgehensweisen zur Bewertung von Queries diskutiert, von denen das Boole'sche genau eines ist. Dass diese Welt komplexer Datenstrukturen in der Kunstgeschichte weitgehend ignoriert wird, ist zweifellos unerfreulich, sollte aber doch wohl kaum der Informatik ganz alleine angelastet werden.

Empfohlene Zitierweise:

Manfred Thaller: Bemerkungen zu kunsthistorischen Informationssystemen; vornehmlich aus der Sicht der Informatik, in: zeitenblicke 2 (2003), Nr. 1 [08.05.2003],

URL: <<http://www.zeitenblicke.historicum.net/2003/01/thaller/index.html>>

Bitte setzen Sie beim Zitieren dieses Beitrags hinter der URL-Angabe in runden Klammern das Datum Ihres letzten Besuchs dieser Online-Adresse. Zum Zitieren einzelner Passagen nutzen Sie bitte die angegebene Absatznummerierung.