

Prof. Dr. Erik Jacobson
Fachhochschule Frankfurt/Main
Fachbereich II (MND)
Kleiststraße 3
D 60 318 Frankfurt

E-mail: ~~jacobson@rz.fh-frankfurt.d400.de~~
~~Tel: 069/1533-2293~~
~~Fax: 069/1533-2262~~
jacobson@fb2.fh-frankfurt.de

den 9.8.1995

Zur Diskussion gestellt:

Ein Basisreferenzmodell zur Informationsverarbeitung

Computer erobern sich immer mehr Einsatzgebiete unserer Welt, angefangen vom Mikrochip in der Waschmaschine, über die Mainframes bei Banken und Versicherungen, bis zu den Millionen PCs, die im Internet zusammengeschlossen sind. Die Beherrschung dieser Instrumente ist Aufgabe der Informatikerinnen und Informatiker, deren Verantwortung mit der Komplexität dieser System wächst. Und bei allen Fachproblemen, denen sie sich hier gegenübersehen, darf jedoch ihr Überblick nicht verloren, sonst kann der eine nicht mehr mit der anderen reden, sonst redet man aneinander vorbei. Hier Hilfestellung zu geben ist Aufgabe der Informatik, die neben den Arbeitswerkzeugen auch Kommunikationsmittel zur Verfügung stellen muß. Eines der wichtigsten hiervon ist ein einheitliches, verbindliches Begriffsgerüst, auf das sich jede und jeder berufen kann, ein Gerüst, das die Grundlage jeder Kommunikation bildet. Für solche Begriffsgerüste haben sich Modelle, Referenzmodelle, Basisreferenzmodelle (BRM) bewährt. Das ISO-OSI-BRM [1] ist eines der schönsten Beispiele hierfür, das auch den Erfolg und die Notwendigkeit von Normungsbemühungen beweist.

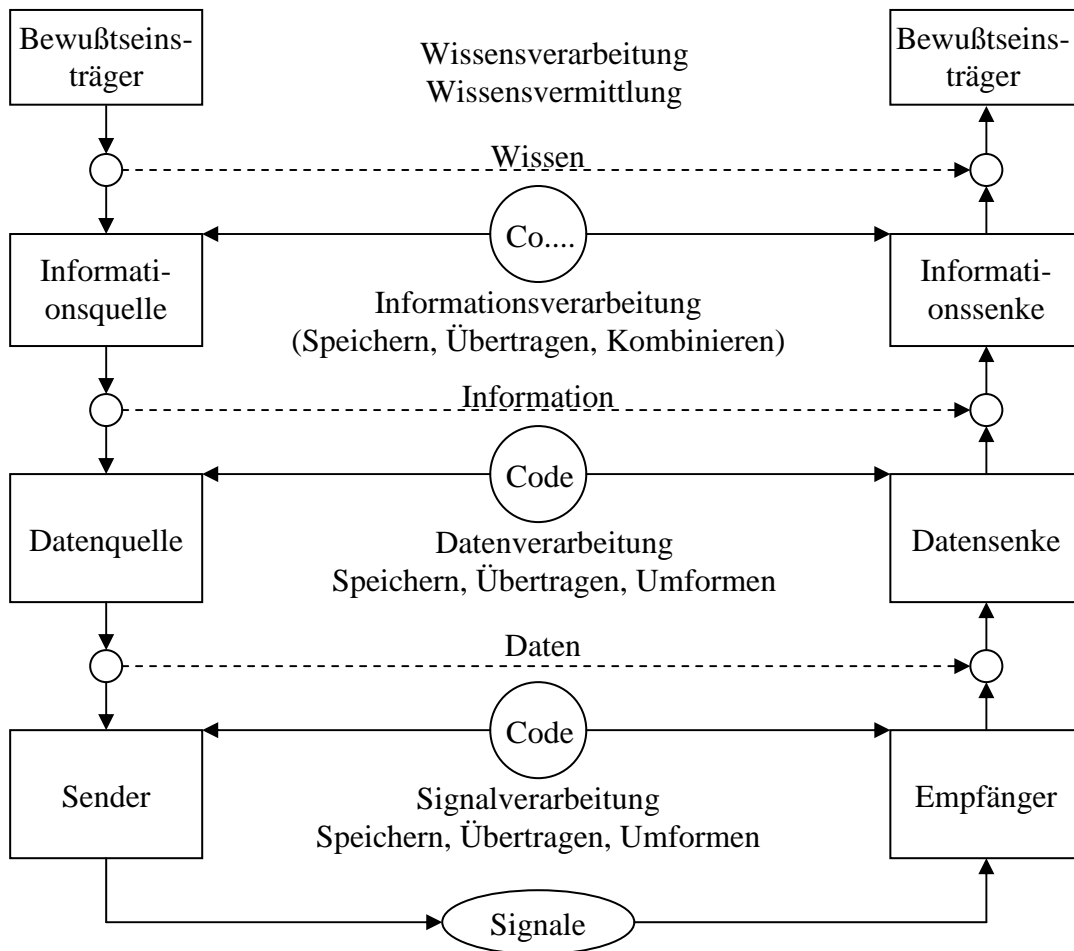
An dieser Stelle soll ein BRM der gesamten Informationsverarbeitung vorgestellt werden, das sich wie ein roter Faden durch die gesamte Literatur zu den Grundlagen der Informatik [2, 3, 4, 5, 6] zieht, das aber in konkreter, formalisierter Darstellung selten zu finden ist. Es geht auf einen Vorschlag von P. Deussen (Univ. Karlsruhe) zurück und ist in zwei Normen bzw. Normentwürfen der ISO und des DIN eingeflossen [7, 8]. Auch die IEC (International Electrotechnical Commission) stützt sich in ihrer Terminologie [9, 10, 11] auf ein solches Modell ab, ohne es explizit zu nennen. Die dort vorgestellten Modelle unterscheiden sich voneinander in einigen Details, auf die hier nicht weiter eingegangen werden soll. Statt dessen soll hier ein weiter auf die relevanten Aspekte reduziertes und um Nuancen angereichertes Modell dargestellt werden, das dieselben Fakten und Formen beinhaltet.

Dieses Modell soll zunächst in seinem formalen Aufbau betrachtet werden:

Als Modellierungsverfahren wird wie in DIN und ISO ein Netz aus Instanzen und Kanälen (NIK) gemäß DIN 66200 gewählt, das den Petrinetzen nahe verwandt ist [12]:

Rechtecke verkörpern Instanzen (Transitionen) oder Verfahren, Kreise verkörpern Kanäle (Stellen) oder Objekte. Beide Knotentypen können nach den bekannten Regeln weiter verfeinert werden. Würde man das hier tun, so ergäben sich weitläufigere Strukturen, die auch zu anderen Modellstrukturen zusammengefaßt werden könnten; womit gesagt werden soll, daß die hier vorgestellte Struktur nur eine von vielen möglichen ist und nicht im Gegensatz zu anderen Modellen stehen muß.

Bild 1:



Topologisch wurde es in die Form eines Schichtenmodells gebracht, wie man es vom OSI-Basisreferenzmodell [1](IS 7498) her kennt. Es umfaßt 4 Schichten:

- Wissen (knowledge),
- Information (information),
- Daten (data),
- Signale (signals).

(Die Signalschicht ist mit der OSI-Schicht 1 - Physical layer - Bitdarstellungsschicht - weitgehend identisch.)

Das BRM zeigt 2 Systeme mit hierarchisch angeordneten Instanzen, die miteinander kommunizieren. Diese Kommunikation erfolgt auf verschiedenen Abstraktionsstufen, wobei die Instanzen einer Schicht auf den Diensten der darunterliegenden Schicht aufsetzen. Bis auf die unterste Schicht erfolgt die Kommunikation virtuell, was mit den gestrichelt gezeichneten Flußpfeilen angedeutet wird. Rein formal kann der gesamte Komplex, der unterhalb einer solchen gestrichelten Flußlinie liegt, zusammen mit den beiden Kanälen (Kreisen) zu einem Kanal (Kreis) zusammengezogen werden, wie es auf der untersten Ebene bei den Signalen angegeben ist.

Zusätzlich werden in diesem Modell noch Kanäle angegeben, welche die Protokolle für diese Kommunikation bereitstellen. Diese Erweiterung scheint deshalb notwendig, weil mit diesem Modell auch das Problem der Protokollbereitstellung, der Synchronisation, des gegenseitigen Verstehens angesprochen werden soll.

Nicht eingezeichnet sind Transitsysteme, die bis in unterschiedliche Abstraktionsschichten hinauf reichen, und wie im OSI-BRM zur Protokollumsetzung bzw. zur Umformung oder Verarbeitung dienen.

Inhaltlich beginnt das Modell in seiner obersten Hierarchiestufe mit der Wissensverarbeitung, die an Bewußtseinsträger, im betrachteten Zusammenhang an Menschen gebunden ist. Hier erfolgt Wissensverarbeitung zum einen sehr lokal in Bewußtseinsträgern (vulgo in den Hirnen), zum anderen findet Wissensvermittlung, Wissenserwerb, Wissensaustausch statt. Dieser Vorgang soll hier vorrangig betrachtet werden.

Nach dem vorliegenden Modell erfolgt dieser Prozeß unter Inanspruchnahme der darunterliegenden Schichten, Maschinen, Prozesse.

Als erster Schritt wird hierbei Wissen in Information umgeformt und abgegeben oder im letzten Schritt zurückgewonnen und aufgenommen. Wie dies geschieht und was dabei alles lokal abläuft, soll hier nicht weiter betrachtet werden. Sicher ist jedoch, daß zu dieser Umformung (gemeinsame) Regeln notwendig sind, die durch Lernprozesse erworben werden, also durch lokale Wissensverarbeitung von aus der Umwelt einfließender Information.

Information wird hier einfach als Objekt betrachtet, in einem Kanal, der den Bewußtseinsträger, den Wissensträger, mit seiner Umwelt verbindet. Auf diese Objekte (Information) können Operationen ausgeführt werden wie Speicherung, Transport (Übertragung) oder Umformungen.

Für eine Verständigung notwendig ist auf jeden Fall ein Kontext oder eine andere Form von Protokoll (Con.?), das vorher bekannt sein muß. Wie es entstanden ist und wie es wirkt, soll hier nicht untersucht werden.

Information hat bei dieser Betrachtung nur eine kurze, ja virtuelle Lebensdauer, denn sie wird sogleich in Daten umgesetzt und diese dann in Signale. Und erst solche existieren in der physikalischen Realität, in Raum und Zeit. In der Informatik wird deren Existenz aber nicht weiter untersucht, sondern erst wieder ihre Zurückwandlung in Daten, Information und Wissen, das beim Empfänger ankommt.

Die Umwandlung der Information in Daten, die Darstellung der Information, erfolgt nach festgelegten Regeln (Codes, s.w.u.) und ebenso die Umwandlung von Daten in Signale. Daß in manchen Systemen die Darstellung in Form von Daten übersprungen werden kann und Information direkt in Signale umgesetzt wird [8], soll hier nur erwähnt werden. Bemerkenswert ist, daß hier für unterschiedliche Umsetzungsprozesse (Information -> Daten, Daten -> Daten, Daten -> Signale) derselbe Begriff, Code, gebraucht wird. Daß dieser Begriff noch weitere Tücken aufweist, wird weiter unten angesprochen.

Ein Beispiel mag den Weg noch verdeutlichen:

Ein elektronischer Brief (eine E-Mail) entsteht zunächst als Gedanke im Kopf, eventuell noch gar nicht sprachlich verfaßt. In Form von Sprache wird er zu Information, dann niedergeschrieben als Text (Daten in Form von Zeichen aus einem Alphabet, als Emotikons aus einer Liste (-:)). Signale entstehen dabei, wenn durch Fingerbewegung die Tasten einer Tastatur in Bewegung gesetzt werden, Ströme fließen, Speicherbits gesetzt (Kondensatoren geladen) werden. Aus den Speicherbits werden wieder Ströme, die im Internet über Drähte fließen, immer wieder in andere Darstellungen umgewandelt werden, bis sie schließlich als Schwärzungen auf Papier fixiert, als optische Signale ins Auge gelangen.

Dieses Beispiel zeigt, wie schwer der Einzelprozeß in der Vertikalen abzugrenzen ist. Wichtig und notwendig ist ein Schema, ein Basisreferenzmodell, das die Betrachtungsebenen eingrenzt.

- Die Signalebene wird hier, im Umfeld der Informatik, nicht nur der Vollständigkeit wegen betrachtet, sondern auch wegen der notwendigen Regeln (Codes) für die Umsetzung der Daten in Signale.

Die gesamte Signalverarbeitung ist Domäne der Naturwissenschaften und Technik, wie Physik und Biologie, Elektrotechnik und Technischer Optik. Hier kann man sich mit geeigneten Vorrichtungen die Signale ansehen, welche die Daten(bits) repräsentieren, die auf einem Magnetband oder in einem Chip gespeichert sind, oder über eine Leitung übertragen werden.

Auf dieser Ebene betrachtet man physikalische Prozesse, die miteinander gekoppelt sind. Diese Ebene könnte man weiter zerlegen und weitere unterlagerte Ebenen finden, z.B. die elementaren Interaktionen von Elektronen, Photonen oder Phononen.

In der Informatik verläßt man sich aber darauf, daß diese Ebene genügend gut modelliert ist und auch funktioniert. Abweichungen von Idealverhalten, also Störungen wie Rauschen, Unterbrechungen oder Verzögerungen, werden dann nur noch phänomenologisch, z.B. in der Informationstheorie [13] auf der Datenebene betrachtet.

- Die Datenverarbeitung ist die klassische Domäne der Informatik. Hierzu soll nicht mehr viel gesagt werden. Im BRM wichtig sind vor allem die Schnittstellen bei der Datenübertragung (Datenkommunikation), die sich einerseits auf die Signalverarbeitung als Basismaschine stützt, andererseits gemeinsame Regeln (Protokolle, Codes) benötigt. Daß in dieser Schicht noch viele Teilschichten (sublayers) eingeführt werden können, ist selbstverständlich. Man denke nur an die Hierarchie der verarbeiteten Datenobjekte, die vom einzelnen Bit über Pixel und Zeichen zu komplexen Dokumenten reicht.

- Wo genau die Datenverarbeitung in die Informationsverarbeitung übergeht, ist heute noch nicht genau festgelegt. Auch hier könnte man Teilschichten einführen, insbesondere wenn man auch dem Computer Formen der Informationsverarbeitung zuschreiben will. Operationen wie Speichern und Übertragen gehören mit Sicherheit dazu; offen ist die Frage, welche Umformungen hierher gehören; ob Veränderungen oder Verlust, ob Filterung oder Kombination mit eingeschlossen werden sollen.

- Noch unschärfer wird die Grenze zwischen Wissen und Information. Hierauf soll hier nicht weiter eingegangen werden. Daß sich hier in den nächsten Jahren noch viel tun wird, verspricht uns das Forschungsgebiet der Künstlichen Intelligenz. Daß wir mit so unscharfen Grenzen umgehen gelernt haben, zeigt uns die Fuzzy-Logik.

Abgesehen von allen bisher aufgezeigten offenen Fragen, für die die Forschung mehr oder weniger schon im Gange ist, zeigt das hier vorgestellte BRM noch eine weitere Klasse von Problemen, die sprachlicher, terminologischer Art sind. Als Prototypen sollen 2 Begriffe herausgegriffen werden:

die Begriffe Signal und Code. Deren Probleme sind zumindest in all den Sprachen gleich, die diese Begriffe als Fremd- oder Lehnwort aus dem Lateinischen benutzen. Beide Begriffe sollen getrennt betrachtet werden, obgleich sich auch Berührungspunkte zeigen:

1. Signal

Folgende Definitionen lassen sich finden:

1. In der Umgangssprache: (Verkehrs)Zeichen [14]
2. In der Fachsprache der Informatik: Darstellung von Information durch den Wert oder Wertverlauf einer physikalischen Größe. [15]

Die Darstellung von Nachrichten oder Daten mit physikalischen Mitteln (DIN 44300 T2)[16].

3. Im IEV (International Electrotechnical Vocabulary) der IEC (International Electrotechnical Commission) IEC 50(702) [10] # 702-000-004-001 signal: A physical phenomenon one or more of whose characteristics may vary to represent information. Note.- The physical phenomenon may be for instance an electromagnetic wave or an acoustic wave and the characteristic may be an electrical field, a voltage or sound pressure.

4. Im ITV (Information Technology Vocabulary) der ISO (International Organisation for Standardization) IS 2382 part 01 "Fundamental terms" [7] # 01.02.01 signal : a variation of a physical quantity used to represent data.

5. Im Entwurf des US Federal Standard (FED_STD-1037C) Glossary of Telecommunication Terms [18] gibt es gleich 4 Definitionen:

- 5.1. Detectable transmitted energy that can be used to carry information.
- 5.2. A time-dependent variation of a characteristic of a physical phenomenon, used to convey information.
- 5.3. As applied to electronics, any transmitted electrical impulse.
- 5.4. Operationally, a type of message,...., with prearranged meanings and which is conveyed or transmitted by visual, acoustic, or electrical means.

6. In englisch-amerikanischen und deutschen Lehrbüchern der Informatik wird "Signal" synonym mit codierter Information (entsprechend der 4. Definition des US FED- STD 1037C) benutzt [18],

7. Die meisten Lehrbücher der Informatik und der Elektrotechnik behandeln das "Signal" als eine Funktion der Zeit "s(t)" ohne viel über deren physikalische Realität zu reden, d.h. sie betrachten mehr das Modell des Objekts als das Objekt selber.[19]

Im Ganzen gesehen hat man tatsächlich zwei verschiedene Begriffe, die mit derselben Bezeichnung belegt sind:

1. Ein Signal ist also zum einen eine konkrete physikalische Entität (ob es nun das physikalische Phänomen ist oder die damit verbundene(n) physikalische(n) Größe(n), sei dahingestellt), ein Objekt, das bearbeitet, verändert, transportiert, gespeichert werden kann, z.B. der elektrische Strom(fluß).

2. Der Begriff Signal kann aber auch auf einem viel höheren Abstraktionsniveau auftreten, nämlich als Information: "Ein Signal geben (zum Start einer Aktion)", ein Ereignis auslösen, eine Meldung, ja eine ganze Nachricht darstellen (".. und winken mit den Äugelein und treten auf den Fuß.."). Die Bearbeitung solcher Objekte wird als Signalisierung bezeichnet, z.B. Signalisierungssystem Nr. 7 (SS7).

- Die Unterscheidung beider Begriffe wird durch den alltäglichen Gebrauch nicht klarer, und auch der Kontext kann hier manchmal nicht helfen, sondern verwirrt noch. Als Beispiel mag das in diesem Zusammenhang oft zitierte Signal an der Eisenbahnstrecke dienen, das dem Zugführer "Halt" oder "freie Fahrt" signalisiert. Es ist zwar ein sehr konkretes physikalisches Objekt, es signalisiert aber auf sehr hohem Abstraktionsniveau nur dem eingeweihten Eisenbahner etwas, nämlich eine Information. Das physikalische Signal, das Phänomen, ist die unterschiedliche räumliche Anordnung eines realen Gegenstands.

Es wird sicher unmöglich sein, hier neue, unterschiedliche Benennungen zu finden, aber wenn man sich dieses Unterschieds bewußt ist, und etwa aus dem Kontext weiß, welcher Begriff gemeint ist, dann wird man damit auch leben können.

2. Code

Beim Codebegriff kann man ebenfalls mindestens zwei unterschiedliche Bedeutungen finden.

Einige Definitionen für "Code":

1. In der Umgangssprache: 1. Zeichensystem als Grundlage für Kommunikation, Nachrichtenübermittlung und Informationsverarbeitung. 2. Schlüssel zu Geheimschriften, Telegraphenschlüssel (Kode). 3. Gesetzbuch (Code Napoleon). [14]
2. In der Fachsprache der Informatik: Eine Vorschrift für die eindeutige Zuordnung von Zeichen eines Zeichenvorrats zu denjenigen eines anderen Zeichenvorrats, DIN 44300 T2 [16]. Der Code mit dem kleinstmöglichen Zeichenvorrat ist der Binärcode mit den zwei Zeichen »0« und »1«. Für die synchrone Datenübertragung werden am häufigsten der ASCII-Code und der EBCDIC eingesetzt [15].
3. Im IEV (International Electrotechnical Vocabulary) der IEC (International Electrotechnical Commission) [9] # 701-000-003-007: Code (F: Code, S: Código): A system of rules defining a one-to-one correspondence between information and its representation by characters, symbols or signal elements. [ebenso CCITT B.13 II.3.7]
4. Im ITV der ISO [21] #04.02.01 code: A collection of rules that maps the elements of one set onto the elements of a second set.
5. Im Entwurf des US Federal Standard (FED_STD-1037C) Glossary of Telecommunication Terms [17] gibt es gleich 6 Definitionen, die sich teilweise überlappen:
 - 5.1 A set of unambiguous rules specifying the manner in which data may be represented in a discrete form.
 - 5.2 [Any] system of communication in which arbitrary groups of letters numbers, or symbols represent units of plain text of varying length.
 - 5.3 A cryptosystem in which the cryptographic equivalents, (usually called "code groups") typically consisting of letters or digits (or both) in otherwise meaningless combinations, are substituted for plain text elements which are primarily words, phrases, or sentences.
 - 5.4 A set of rules that maps the elements of one set, the coded set, onto the elements of another set, the code element set.
 - 5.5 A set of items, such as abbreviations, that represents corresponding members of another set.
 - 5.6 To write a routine.
6. In den einschlägigen Lehrbüchern [20] wird in der Regel der Code als eine Vorschrift für die Abbildung aus einer Menge in eine andere, vorwiegend von Daten in Signale, oder Daten auf Daten gesehen. Die Abbildung von Informationen auf Daten wird hier weniger oft betrachtet.
7. Im Jargon der Datenverarbeiter ist Code auch das Ergebnis einer Programmierarbeit (s.5.6).
8. In der Telekommunikation kann ein Code auch ein Datenobjekt sein, z.B. "postal code" in einer Adresse (X.400) oder der "country code" einer Telefonnummer (Q.10) [22]

Faßt man die gefundenen Definitionen zusammen, so findet man wie beim Begriff "Signal" zwei Bedeutungen:

1. Eine abstrakte Vorschrift für die Abbildung von Objekten, die im BRM in mehreren Schichten gefunden werden kann;
2. das Ergebnis eines solchen Vorgangs in Form von Datenobjekten.

Gegen die letzte Interpretation gibt es viele Bedenken. Wahrscheinlich wird man auch hier mit den 2 Bedeutungen leben müssen.

Insgesamt zeigen diese Betrachtungen, wie inhomogen die Terminologie der Informatik ist und welche Inkonsistenzen sie birgt. Ein Grund dafür mag in der Entstehungsgeschichte der Informatik zu finden sein, denn mehrere Disziplinen trugen dazu bei (Mathematik, Elektrotechnik, Nachrichtentechnik, etc), die wohldefinierte Terminologien benutzen; ihre Vermischung bringt jedoch ein Chaos.

Leider legt sich dieses Problem nicht, sondern mit der Einführung neuer Begriffe, die oft aus Marketinggründen aus der Rolle fallen (sollen), verschlimmert sich die Situation weiter.

Der vorliegende Beitrag sollte diese Problematik aufzeigen und einen Ansatz zur Stabilisierung zur Diskussion stellen.

Literatur:

- 1: IS 7498: Open Systems Interconnection - Basic Reference Model.
- 2: Bauer/Goos: Informatik 1. Springer 1991. ISBN 3-540-52790-7
- 3: Schumny: Digitale Datenverarbeitung. Vieweg 1989 ISBN 3-528-14031-3
- 4: Schaback: Grundlagen der Informatik. Vieweg, 1988 ISBN 3-528-06304-1
- 5: Koch/Rembold: Einführung in die Informatik. Hanser, 1977. ISBN 3-446-12385-7
- 6: Dworatschek: Grundlagen der Datenverarbeitung. W.de Gruyter, 1986. ISBN 3-11-010631-0
- 7: IS 2382 part 01: Information Technology Vocabulary - Fundamental terms.
- 8: E DIN 44 300, Teil 1: Informationsverarbeitung - Allgemeine Begriffe.
- 9: IEC 50(701): International Electrotechnical Vocabulary - Telekommunikation
- 10: IEC 50(702): International Electrotechnical Vocabulary - Schwingungen, Signale und zugehörige Einrichtungen
- 11: IEC 50(704): International Electrotechnical Vocabulary - Nachrichtenübertragungstechnik
- 12: Reisig W.: Systementwurf mit Netzen. Springer 1985 ISBN: 3-540-13786-6
- 13: DIN 44301: Informationstheorie - Begriffe. (Nov.1984)
- 14: Duden - Das Fremdwörterbuch. 4. Auflage, Dudenverlag. ISBN 3-411-20905-4
- 15: Lipinski, K. (Hrsg.): Lexikon der Datenkommunikation. Datacom 1995. ISBN: 3-89238-101-1
- 16: DIN 44300 Teil 2: Informationsverarbeitung - Begriffe - Informationsdarstellung (Nov.1988)
- 17: NTIA-ITS: US Federal Standard (FED_STD-1037C) Glossary of Telecommunication Terms
- 18: Tanenbaum A.S.: Operating Systems. Prentice-Hall 1987
- 19: Johnson J.R.: Digitale Signalverarbeitung. Hanser 1991 ISBN: 3-446-15890-1
 Lochmann, D.: Digitale Nachrichtentechnik. Verlag VEB Technik 1990 ISBN: 3-341-00699-0
 Mildenerger, O.: System- und Signaltheorie. Vieweg 1989 ISBN: 3-528-13039-3
 Oppenheim/ Willsky: Signale und Systeme. Verlag: VCH 1989 ISBN: 3-527-26712-3
 Russer P.: Informationstechnik. Verlag: VCH 1988 ISBN: 3-527-26934-7
 Schrüfer, E.: Signalverarbeitung. Hanser 1992 ISBN: 3-446-16563-0
- 20: Hamming, R.W.: Information und Codierung. VCH 1987 ISBN: 3-527-26611-9
 Mildenerger, O.: Informationstheorie und Codierung. Vieweg 1990 ISBN: 3-528-03046-1
 Sweeney,P.: Codierung zur Fehlererkennung und -korrektur.Hanser 1992 ISBN:3-446-16439-1
- 21: IS 2382 part 04: Information Technology Vocabulary - Organization of data
- 22: ITU-T / CCITT-Recommendations 1993; InfoMagic-CD-ROM "Standards".