

Forschungs- und Entwicklungsprojekte

Informatik Forsch. Entw. (2004) 18: 88–93
Digital Object Identifier (DOI) 10.1007/s00450-003-0140-6

In dieser Rubrik erscheinen in unregelmäßiger Folge Kurzdarstellungen geplanter, laufender oder abgeschlossener Projekte. Die Darstellungen werden in der Regel von den Projektbeteiligten geliefert. Die Auswahl erfolgt durch die Herausgeber. Dabei wird die Bedeutung des Projekts für die Fortentwicklung der Informatik das Hauptkriterium sein. Bei geplanten und laufenden Projekten ist ein wichtiges Kriterium der Wunsch, Kontakte zu etablieren und die Zusammenarbeit zwischen verschiedenen Gruppen zu fördern. Bei abgeschlossenen Projekten geht es primär um die Vermittlung von Erfahrungen und Ergebnissen, die sich nicht für die Veröffentlichung in redaktionellen Beiträgen eignen.

Der Informationsbegriff – eine informatikorientierte Annäherung

Albert Endres

Sindelfingen (e-mail: a.endres@acm.org)

Zusammenfassung. Information ist der wichtigste Grundbegriff der Informatik. Ein semantischer Ansatz erklärt Information als Bedeutung einer Nachricht für ihren Empfänger. Als Folge davon hängt der Informationsgehalt einer Datei, einer Datenbank, eines Buches oder eines Filmes von der Interpretation ab, die ein Nutzer vornehmen kann oder gerade vornimmt. Informationsmaße werden sowohl von der Lesbarkeit einer Nachricht als auch von der Qualität des Interpretieren bestimmt. Information kann neu entstehen, verloren gehen oder verändert werden.

Schlüsselwörter: Information, Informationstheorie, Semantik

Abstract. Information is the basic concept of computer and information science. A semantic approach defines the information content of a message by its meaning for a recipient. As a consequence, the information content of a file, a database, a book or a video depends on the interpretation that a user is able to perform or is just performing. An information metric is determined both by the readability of an input message and the power of the interpreter. Information can be generated, lost or modified.

Keywords: information, information theory, semantics

CR Subject Classification: E.4, H.1

1 Einleitung

Informatik wird meist beschrieben als die Wissenschaft, Technik und Anwendung der maschinellen Verarbeitung von Information. Bezüglich einer Definition des Begriffs Information gibt es leider bis heute keine einheitliche Meinung. In einem Buch aus dem Jahre 1991 gibt Zemanek [10] zehn verschiedene, allesamt nicht befriedigende Definitionen an. Auch der Duden Informatik [1] stellt fest, dass wir von einer Klärung des Begriffs Information noch weit entfernt sind. In aktuellen Beiträgen im Informatik-Spektrum weisen sowohl Klemm [6] wie Rechenberg [7] auf schädliche Folgen hin, die sich aus der

fehlenden oder falschen Behandlung dieses Themas für die Informatik ergeben. In dieser für die Informatik so misslichen Situation riskiere ich einen Anlauf. Zwecks Begründung und Illustration verweise ich auf historische Bezüge und greife auf einige Beispiele zurück. Vielleicht trägt der Vorschlag auch etwas zum Selbstverständnis der Informatik bei.

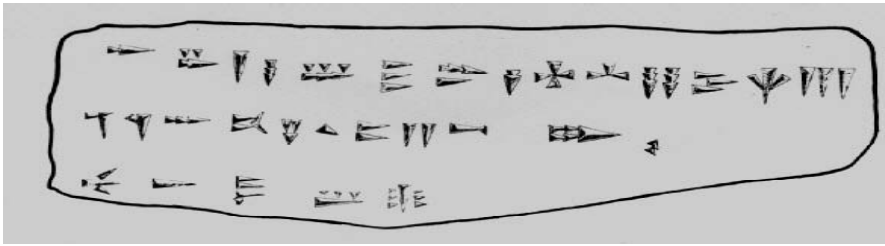
2 Kurze Geschichte der menschlichen Kommunikation

Am Anfang jeder menschlichen Kommunikation stehen Zeichen und Töne. Das gilt sowohl für die Menschheit als Ganzes als auch für jedes neugeborene Kind. Gesichtsausdruck, Hand- oder Armbewegungen, Körperhaltungen und unartikulierte Laute stehen als erstes zur Verfügung. Dann können Gegenstände dazu kommen wie Essgeschirr, Spielzeug, Steine, Tücher, Buschtrommeln oder Rauchzeichen. Die gesprochene Sprache kommt erheblich später. Ganz zuletzt kommt die Schrift. Hier gibt es die drei Grundarten, die auf Bildern, auf Silben oder auf Lauten basieren, sowie deren Kombinationen. Die ältesten Schriften sind Bilderschriften, da sie ein geringeres Maß an Abstraktion verlangen als Silben- oder Lautschriften. Aus der Zeit um 1400 vor Christus stammt die heutige chinesische Schrift [9]. Sie ist eine Bilderschrift und benötigt so viele Bilder wie es Begriffe gibt (etwa 100.000). Die Schrift der Mayas in Mittelamerika war im Grunde eine Silbenschrift. Sie umfasste etwa 300 Zeichen [8].

Eine auf unsere Kultur nachhaltig wirkende Lautschrift wurde um 1400 vor Christus von Phöniziern in Ugarit (heute Ras Shamra) im Norden des heutigen Syrien erfunden. Es handelte sich um ein Keilschrift-Alphabet mit 29 Zeichen (siehe Abb. 1). Die Schrift wurde aus einer älteren Silbenschrift abgeleitet, in der die ersten drei Zeichen für Aleph (Ochse), Beit (Haus) und Gimmel (Kamel) standen. Obwohl die den Phöniziern folgenden Griechen keine Keilschrift mehr schrieben, nannten sie die Zeichen ihrer Lautschrift weiterhin Alphabet.

3 Information in Biologie und Technik

Die biologische Welt ist vor allem dadurch gekennzeichnet, dass sie große Mengen an Information generiert, speichert

Abb. 1. Alphabet von Ugarit¹

und überträgt. Information ist nicht an Menschen gebunden. Sie spielte längst eine zentrale Rolle, bevor es Menschen gab. Die Pflanze, die sich dem Licht zuwendet, die Biene, die mit einem Tanz den Weg zum Rapsfeld beschreibt, oder der Vogel, der seine Route von Skandinavien nach Afrika findet, sie alle verarbeiten Information. Ob und wie weit Information auch in der nicht-belebten Natur, also bei Steinen und Metallen, am Werke ist, darüber wird noch kräftig spekuliert. Der bei Lebewesen durch Vererbung zu übertragende Teil an Information wird in Form von DNA-Basen codiert. Sie steuern die Fortentwicklung der Art (Phylogenese). Man kann den DNA-Code als Alphabet auffassen, bestehend aus vier Zeichen. Ihre Anordnung im menschlichen Erbgut, aber noch nicht ihre Bedeutung, d.h. ihre Wirkungen, hat man mittlerweile entschlüsselt. Das von einem Individuum während seines Lebens erlernte Wissen wird ebenfalls als Information gespeichert (Ontogenese). Beim Menschen geschieht dies größtenteils im Langzeitgedächtnis des Gehirns. Die Codierung ist noch nicht genau verstanden.

In der Technik gibt es eine Vielzahl von beobachtbaren Signalen. Manche Formen können von Menschen nur mit Hilfe spezieller Geräte registriert werden. So misst ein Tachometer Entfernungen und Geschwindigkeiten, ein Thermometer Temperaturen und ein Geigerzähler die Radioaktivität. Die Position eines Zahnrads oder die Höhe einer Quecksilbersäule geben zunächst in analoger Form Auskunft über den jeweiligen Wert. Steht die Speicherung und Verarbeitung von Signalen im Vordergrund, sprechen wir von Daten. Geht es primär um Kommunikation, so benutzen wir den übergeordneten Begriff Nachricht. Daten und Nachrichten können analog oder digital codiert werden. In Rechnern werden Daten nur noch digital gespeichert, sehr oft in Form einer Binärcodierung. Sie werden bei der Eingabe von einer externen Form in die interne Form umgewandelt und bei der Ausgabe wieder als sichtbare Zeichen oder als hörbare Töne dargestellt.

Der Datentyp, der zurzeit am häufigsten in Rechnern gespeichert und bearbeitet wird, sind Texte. Nur in der westlichen Welt sind Texte als alphabetische Zeichenketten codiert, die von links nach rechts gelesen werden. Im chinesischen, koreanischen und japanischen Sprachraum haben sich eine vereinfachte Form einer Bildschrift (Kanji) sowie mehrere Silbenschriften (Katakana, Hirakana) durchgesetzt. Abbildung 2 zeigt einen Werbetext der japanischen Fernsehgesellschaft NHK, in dem Schriftzeichen aus vier Schriften vorkommen, nämlich Kanji, Katakana, Hirakana und Latein [5]. Dies entspricht der typischen Situation im japanischen Geschäftsleben und im Straßenbild japanischer Städte. In China und Japan

bietet selbst der kleinste Palmtop heute direkte Ein- und Ausgabemöglichkeiten für Tausende von Zeichen.

4 Numerisches Beispiel

Um die Spannweite des Informationsbegriffs zu illustrieren, sollen zunächst drei Beispiele gegeben werden. Das erste basiert auf dem Zahlbegriff. Um den uns geläufigen Assoziationen vorzubeugen, benutze ich römische Zahlen. Gegeben sei ein Ausdruck

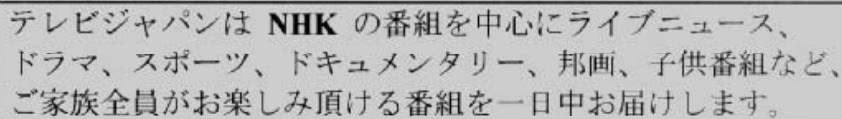
„IV plus VII“

Der Benutzer dieser Schreibweise wusste, dass man beide Hände zu Hilfe nehmen darf. Ein „V“ steht für ein ganze Hand. Steht ein Finger links davon, wird er abgezogen, steht er rechts, wird er hinzugefügt. Das lateinische Wort „plus“ heißt „mehr“. Wir interpretieren es hier als Aufforderung zwei Zahlen zu addieren. Indem man „IV“ umwandelt in „IIII“, ergibt die Addition zunächst „VIIIII“. Dies ist unökonomisch und fehleranfällig und lässt sich umwandeln in „VVI“ oder „XI“, wobei „X“ für zwei Hände steht. Die in dem Ausdruck vorkommenden Zeichen werden quasi auf die menschlichen Hände abgebildet, um sie zu interpretieren.

5 Wortbasiertes Beispiel

Das nächste Beispiel basiert auf kurzen alphabetischen Zeichenketten oder Wörtern. Die möglichen Kombinationen der zwei ersten Buchstaben unseres Alphabets in Groß- und Kleinschreibung ergeben 20 verschiedene Zeichenketten. Diese können je nach Kontext verschieden interpretiert werden. In Tabelle 1 sind in den sechs rechten Spalten sechs Möglichkeiten der Interpretation angegeben. Nicht betrachtet wurden andere als die heutige deutsche Hochsprache (Althochdeutsch, Mundarten) und andere Kulturkreise (Fremdsprachen, Jargon), ebenso wenig Moody's Ratings, Schulnoten, Bodenbewertungen und dgl. In einigen Fällen sind alle sechs rechten Spalten leer, d.h. es gibt keine gültige Interpretation. Die entsprechende Zeichenfolge enthält keinerlei Information für einen der sechs Kontexte. (Man kann argumentieren, dass sie für jemanden, der nur Schreibübungen macht, schon Information enthält. Das wäre ein anderer Kontext.) Bei den Abkürzungen sind einige Zeichenketten mehrfach belegt. Um sie richtig zu interpretieren, braucht man eine Verfeinerung des Kontexts.

¹ Diese Skizze ist in einer Tontafel aus dem Museum von Aleppo nachempfunden (von H. Zaroura).



テレビ日本は **NHK** の番組を中心にライブニュース、
ドラマ、スポーツ、ドキュメンタリー、邦画、子供番組など、
ご家族全員がお楽しみ頂ける番組を一日中お届けします。

Abb. 2. Aktueller japanischer Werbetext²

6 Satzbasierendes Beispiel

Aus Zeichenketten oder Wörtern werden Sätze gebildet. Auf den folgenden Satz brachte mich ein Fachartikel [4]:

„Green destiny saves power“.

Eine naheliegende deutsche Übersetzung unter Berücksichtigung lexikalischer und grammatikalischer Regeln wäre „Das grüne Schicksal rettet die Herrschaft“. Um den wahren Gehalt dieses Satzes abzuleiten, sollte man wissen, dass es sich bei „Green Destiny“ um die Kennzeichnung eines Rechner-Prototypen handelt und dass mit „power“ Strom gemeint ist. Auch hier ist die korrekte Interpretation nur mit Hilfe eines speziellen Kontexts möglich. Eine sinnvollere Übersetzung wäre daher „Green Destiny spart Strom“. Genau diese Differenzierung wird von einer heute handelsüblichen maschinellen Übersetzung verlangt.

Auch wenn wir den Regeln einer Grammatik folgen, können wir Sätze bilden, deren Sinn uns entgeht. In dem Satz „Rote Krähen dösen geil“ enthält zwar jedes Wort Information. Dass der Satz als Ganzer unsinnig ist, stellt sich erst heraus, wenn wir mehr Information über die Bedeutung der einzelnen Wörter heranziehen. Dasselbe Problem kann sich eine Stufe darüber wieder ergeben, wenn wir mehrere Sätze zusammen betrachten. Ja, es setzt sich beliebig weiter nach oben fort. Es führt dies zur Grundproblematik derjenigen Information, mit der wir Wissen darstellen, dass nämlich ein Weltverständnis dahintersteckt. Es darf dies jedoch kein Grund sein, sich mit einem rein syntaktischen oder gar nachrichtentechnischen (d.h. Shannons) Informationsbegriff zufrieden zu geben.

7 Definition des Begriffs Information

Mit diesen Beispielen nähern wir uns einer Begriffsdefinition. Ein *Informationsraum* ist ein Tripel

$$I = (Z^*, S, K),$$

wobei $Z^* = \{u_1, u_2, \dots, u_n\}$ eine Menge von Wörtern u_i über einem Alphabet Z , S eine Menge von Symbolen und K einen Kontext bezeichnet. Die Wörter können beliebige Signalfolgen oder Nachrichten sein, die entweder analog oder digital codiert sind. Folgen können sich sowohl in zeitlicher wie in räumlicher Anordnung befinden. Um eine Nachricht zu interpretieren, wird sie nach vorher vereinbarten Regeln zerlegt. Bei Sprachen heißt diese Struktur Syntax³, bei Schachpartien sind es Stellungen. Die Bedeutung oder (semantische)

² Auf deutsch: NHK bietet tägliches Ganztags-Programm; von den neuesten Nachrichten über Drama, Sport und Dokumentarfilme bis zu Kinderprogrammen

³ Nicht einmal bei kontextfreien Sprachen kann man ihre Semantik allein aus deren Worten, d.h. den Terminalsymbolen, ableiten. Dass Nicht-Terminale oft ebenso wichtig sind, ist hier vernachlässigt.

Information I eines solchen Wortes u lässt sich als partielle Abbildung auffassen, und zwar

$$s = I(u, k) \quad \text{oder} \quad I : Z^* \times K \rightarrow S$$

Die Symbole S stammen immer aus einem für die Interpretation zugänglichen oder dem Interpreten bekannten Vorrat. Meist stehen die Symbole für Objekte der realen Welt. Bei einem menschlichen Interpreten können die Symbole abstrakt sein, d.h. es muss ihnen kein reales Objekt entsprechen. Es können nur in der Vorstellung existierende Konzepte oder auch Gefühle sein. Bei einer Interpretation durch Rechner müssen die Symbole ihrerseits als Zeichenketten oder Operationsfolgen codiert sein. Ein Beispiel ist ein Lexikon, das die Vokabeln der Zielsprache enthält, die den Vokabeln der Quellsprache zugeordnet sind. Bei einer Programmiersprache können die Symbole aus kleinen Codestücken oder Codegerippen bestehen, mit welchen die Bedeutung quellsprachlicher Ausdrücke nachempfunden wird. Je mehr Symbole zur Verfügung stehen, je häufiger sind wir in der Lage, eine Nachricht zu interpretieren und umso differenzierter tun wir dies. Wenn man von einer fremden Sprache nur drei Worte zuordnen kann, geht der Rest an uns vorbei. Es findet keine Kommunikation statt.

Auch Symbole oder Objekte können miteinander verknüpft sein. Wir sprechen dann von Strukturen, Modellen oder Theorien. (Diese Begriffe hier näher zu erläutern, würde zu weit führen.) Eine häufige Form sind Aussagen (Propositionen) über Eigenschaften von Objekten. Bei Kleinkindern beginnt das mit „Baby hungrig“ oder „Windeln nass“, wobei Mütter diese Information auch im vorsprachlichen Alter des Kindes bereits wahrnehmen können. Später folgt mit „Ich mit Vati Auto fahren“ eine Verknüpfung von mehreren Objekten.

Der Kontext K kann entweder explizit vorgegeben sein, oder aber er wird während der Interpretation abgeleitet. In diesem Falle wird die Interpretation in einem allgemeineren Kontext begonnen, wird dann aber unterbrochen, um zusätzliche Kontextinformation zu gewinnen und einzubeziehen. Es ist hier angenommen, dass der jeweilige Kontext durch eine Zeichenkette identifiziert wird. Sehr oft hängt er von vielen Faktoren ab und muss erst berechnet werden. Die Abbildung I kann auf verschiedene Arten angegeben werden. Mal hat sie Tabellenform, mal ist es ein Algorithmus. Es kann aber auch eine durch Erlernen erworbene nicht-algorithmische Beziehung sein, die sich durch neuronale Netze beschreiben lässt. Die Abbildung ist partiell in dem Sinne, dass sie nicht für alle Punkte des Eingabebereichs definiert ist. Nicht alle möglichen Zeichen- oder Tonkombinationen sind brauchbar (wie bei dem Beispiel in Tabelle 1).

8 Informationsmaße

Ob und wie man Information messen kann, möchte ich nur andeutungsweise behandeln. Um sinnvoll zu sein, muss ein

Tabelle 1. Buchstabenfolgen und ihre Interpretationen

Zeichenfolge	Deutsche Sprache	Abkürzung	Autokennzeichen	Maßeinheit	Tonbezeichnung	Chemisches Zeichen
a		Anno, Atto		Ar	a-Moll	
aa						
aA						
ab	weg, fort (nicht Satzanfang)					
aB						
A		Anfang (A bis Z, A und O)	Augsburg	Ampere	A-Dur	
Aa	Kot (Kindersprache)					
AA		Auswärtiges Amt, Anonyme Alkoholiker	Aalen			
Ab	weg, fort (Satzanfang)					
AB		Autobahn	Aschaffenburg			
b					b-Moll	
ba						
bA						
bb						
bB						
B		Brief (auf Kurszetteln), Bundesstraße	Berlin	Bel, Byte	B-Dur	Bor
Ba						Barium
BA		Bachelor of Arts, British Airways	Bamberg			
Bb						
BB		Brigitte Bardot	Böblingen			

elementares Maß ausdrücken, wie weit eine Nachricht interpretiert werden kann. Vorstellen muss man sich dafür ein Spracherkennungsprogramm, das Tonfolgen analysiert, oder einen Übersetzer, der von einer natürlichen Sprache in eine andere übersetzt. Der Erfolg dieser Programme hängt von zwei Faktoren ab, nämlich der Lesbarkeit der Eingabe und der Qualität des Programms. Anders ausgedrückt, bei gleichem Programm kann eine bestimmte Eingabenachricht besser oder schlechter zu interpretieren sein als eine andere. Entweder gibt es viele Störgeräusche und Tippfehler oder wenige. Bei gleicher Eingabe kann ein neueres Programm erfolgreicher sein als ein älteres. Eine weitere Überlegung betrifft die Frage, wie syntaktische Einheiten gezählt werden. Nimmt man in Tabelle 1 als Einheit die einzelnen Zeichenketten, besteht hier die Eingabe aus 20 Einheiten. Zu beachten ist, dass im Gegensatz zu diesem Beispiel Information in der Regel nicht aus isolierten Partikeln besteht, die man einfach abzählt, sondern aus Strukturen. Diese bauen aufeinander auf und können einander vertreten. Bei Programmiersprachen handelt es sich dabei um Unterprogramme, Anweisungen, Operatoren und Operanden. Ein Maß könnte ausdrücken, welchen Prozentsatz von Übungsprogrammen ein bestimmter Compiler versteht oder ein Rechner ausführen kann.

Schließlich kann man noch nach Kontexten differenzieren. Auch hierfür sei Tabelle 1 herangezogen. Von den sechs angegebenen Kontexten kommt die letzte Spalte am schlechtesten weg. Es können nur zwei von 20 Buchstabenfolgen den Namen von chemischen Elementen zugeordnet werden. Ist M_k

das Informationsmaß eines Kontext k , dann ist $M_6 = 10\%$. Am besten ist die Spalte Abkürzungen mit sieben Treffern ($M_2 = 35\%$). Als Maß M über alle sechs Kontexte K ergibt sich $M = 12/20 = 60\%$.

Ist N die Anzahl der Eingabewörter, lässt sich das Maß M_k wie folgt ausdrücken:

$$M_k = \left(\sum_{n=1}^N (\text{if } \underline{def} I(u_n, k) \rightarrow 1, \underline{else} 0) \right) / N.$$

Darin ist \underline{def} eine Boolesche Funktion, die feststellt, ob die Abbildung \bar{I} definiert ist, d.h. ob beide Seiten vorhanden sind. Für das zweite Maß M eignet sich die Form eines Algorithmus besser:

```

M := 0 ;
for n = 1 to N do MM[n] = 0 enddo ;
for n = 1 to N do
  for k = 1 to K do
    if def I(u_n, k) then MM[n] := 1 endif
  enddo
enddo ;
for n = 1 to N do M := M + MM[n] enddo ;
M := M/N

```

Die genauere Definition dieser Maße wie auch die Komplexität ihrer Berechnung hängen davon ab, wie syntaktische Einheiten gezählt bzw. von einem Parser erkannt werden. Mit

Hilfe dieser elementaren Maße lassen sich bereits Fragen der Art behandeln, wie sie im Duden Informatik [1] unter dem Stichwort Information gestellt sind. Etwa, durch welche Operationen vergrößert oder verringert sich der Informationsgehalt einer Nachricht? Ähnlich wie in der Softwaretechnik sind noch viele andere Maße vorstellbar, die nicht nur aussagekräftiger, sondern auch nützlicher sind.

9 Entstehung, Verlust und Änderung von Information

Bezogen auf einen einzelnen Empfänger, aber auch auf alle Lebewesen, kann Information neu entstehen, verloren gehen oder verändert werden. Es bedarf dazu einer neuen Vereinbarung zwischen den Beteiligten. Bei Menschen kann dies bewusst oder auch unbewusst geschehen. Zu letzterem gehört das Vergessen. Vergleichen wir Kinder unterschiedlicher Kulturkreise, so sind es nur ganz wenige Laute und Zeichen, die sie bereits im Moment der Geburt auf die gleiche Weise interpretieren. Den allergrößten Teil lernen sie von anderen Menschen während ihrer ersten Lebensjahre. So lernt jedes chinesische und japanische Schulkind über Jahre hinweg, Verknüpfungen zwischen Tausenden von Bildzeichen einerseits und seiner realen oder geistigen Umwelt andererseits herzustellen. Fahrschüler lernen Verkehrszeichen und Elektriker lernen, wie man Schaltpläne liest. Die Vereinbarungen, die hinter all diesen Sprachen stehen, werden nicht von den Nutzern selbst getroffen. Anders ist es bei einem Freundeskreis oder einer Sportmannschaft. Sie können eine eigene Zeichensprache entwickeln, die sie vielleicht sogar geheim halten. Ein Extremfall sind Abkürzungen. Manchmal werden sie für die Zwecke eines einzelnen Fachartikels oder Vortrags eingeführt und anschließend wieder vergessen. In maschinellen Systemen können sehr leicht Zuordnungen, wie sie in Tabelle 1 angedeutet sind, hinzugefügt, gelöscht oder verändert werden.

Aufgrund des besonderen Lernprozesses in der frühen Kindheit und dem damit verbundenen Vergessen (kindliche Amnesie) ist uns Menschen nicht genau bewusst, über welche Interpretationsmöglichkeiten wir eigentlich verfügen. Sie mögen uns manchmal fast unbegrenzt vorkommen. Oft sind wir jedoch überrascht, wie schnell wir an Grenzen stoßen und welche Fehler wir machen. Bei maschinellen Systemen sind die Verhältnisse im Vergleich dazu meist recht übersichtlich.

10 Was folgt daraus für die Informatik?

Der Informationsgehalt einer Zeichenkette, einer Grafik, einer Bilder- oder Tonfolge, eines Buches, eines Films oder einer Datenbank ist weder inhärent noch konstant. Es bedarf einer Vereinbarung zwischen Ersteller und Nutzer, oder zwischen Sender und Empfänger, damit der Sinn verstanden wird. Diese Vereinbarung kann implizit erfolgen, wenn man lange genug in derselben Gruppe oder im selben Kulturkreis gelebt hat. Man benutzt dann einen gemeinsamen Symbolvorrat und kann den Kontext richtig einschätzen oder leicht herbeiführen. Die Information kann sich ändern, wenn neue Interpretationen dazu kommen. Diese können beim Nutzer entstehen, ohne dass der Ersteller dies weiß oder einen Einfluss darauf hat. Sie können vor der Erstellung, während derselben oder danach entstehen.

Da die Interpretation vom Interpretieren abhängt, ist sich der Sender einer Nachricht, der Autor eines Buches oder der Ersteller einer Datenbank nie sicher, was der Empfänger oder Nutzer daraus entnimmt. Man kann Annahmen machen bezüglich des Symbolvorrats, über den der Nutzer verfügt, und bezüglich des Kontexts, in dem sich der Nutzer im Moment des Empfangs befindet. Diese Annahmen können aber falsch sein. Es kann passieren, dass der Nutzer die Nachricht oder Zeichenkette anders interpretiert, als dies der Sender oder Ersteller beabsichtigte. Es kann auch sein, dass der Empfänger Information hineinliest, die der Ersteller gar nicht preisgeben wollte, etwa den Zeitpunkt der Erstellung, die gesellschaftliche Position oder politische Grundhaltung des Absenders. Information kann für den Empfänger redundant sein, nämlich dann, wenn er den durch sie vermittelten Tatbestand bereits kennt. Sie kann für ihn sehr wertvoll sein oder auch wertlos, je nachdem in welcher Situation er sich gerade befindet. Dieselbe Information kann wiederholt übertragen werden, entweder mittels derselben Nachricht oder mittels einer den Inhalt wahren Veränderung. Zu letzterem gehört die zweite Übersetzung des obigen englischen Satzes ins Deutsche. Es kann auch sinnvoll sein, inhaltlich gleiche Information mehrfach zu speichern, sei es aus Sicherheitsgründen oder um unterschiedliche Nutzerkreise anzusprechen. Die Übersetzung aller UNO-Beschlüsse oder Protokolle des EU-Parlaments in mehrere Sprachen dient diesem Zweck. Dass sie außerdem in HTML, PDF und Word angeboten werden, hat technik-interne Gründe.

Viele Diskussionen über das Wesen von Information werden dadurch überlagert, dass man gleichzeitig immer auch von ihrem Zweck oder ihrer Verwendung spricht. Beispiele dafür sind Unterhaltung, Werbung, Indoktrination oder Bildung. Für eine Diskussion des sehr wichtigen Zusammenhangs von Information und Wissen sei auf Endres [2] oder Endres/Fellner [3] verwiesen.

Die Informatik tut sich keinen Gefallen, wenn sie der Frage, was Information ist, aus dem Wege geht. Sie sollte sich nicht in Datentechnik oder Datenverarbeitung umbenennen, sondern sich sowohl in theoretischer wie in praktischer Hinsicht um die maschinelle Informationsverarbeitung kümmern, wie sie sich aus dem hier benutzten (semantischen) Informationsbegriff ergibt. Bei Texten setzt diese seit über 3000 Jahren oberhalb des Niveaus von Schriftzeichen ein, seien es phönizische oder chinesische. Bei dieser Betrachtungsweise erscheinen nicht nur Programmiersprachen, Datenbanken und Betriebssysteme, sondern auch Gerätetechnik (Bildschirme, Drucker) und Rechnerarchitekturen in einem neuen, stärker fokussierten Licht. Wenn es um Informationsverarbeitung geht, sollten wir uns primär in Richtung auf den menschlichen Nutzer hin orientieren. Wenn wir Informatiker das nicht tun, wer denn sonst?

Danksagung. Die in diesem Beitrag wiedergegebenen Ideen sind das Ergebnis einer über Jahre sich erstreckenden Diskussion mit interessierten Fachkollegen. Danken möchte ich insbesondere Volker Claus, Peter Rechenberg, Norbert Wartig und Heinz Zemanek, sowie den beiden Hauptherausgebern dieser Zeitschrift, die bei der Entstehung dieses Beitrags als Resonanzboden dienten.

Literatur

1. Duden Informatik, 3. Auflage (2001) Claus V und Schwill A (Bearb.) Dudenverlag, Mannheim
2. Endres A (2003) Die Wissensgesellschaft und ihr Bezug zur Informatik. Informatik Spektrum 26(3): 195–200
3. Endres A, Fellner DW (2000) Digitale Bibliotheken. Informatik-Lösungen für globale Wissensmärkte. dpunkt, Heidelberg
4. Feng W (2003) Making a case for efficient supercomputing. ACM Queue 1(7): 54–65
5. Hensch K (2003) History of Eastern Languages in Computing. Research and Development in IBM. Eigenverlag
6. Klemm H (2003) Ein großes Elend. Informatik-Spektrum 26(4): 267–273
7. Rechenberg P (2003) Zum Informationsbegriff der Informationstheorie. Informatik-Spektrum 26(5): 317–326
8. Schele L, Freidel D (1990) Die unbekannte Welt der Maya. Knaus, München
9. Wang W (2000) Die chinesische Sprache. Spektrum der Wissenschaft, Dossier: Sprachen: 72–78
10. Zemanek H (1991) Weltmacht Computer – Weltreich der Information. Bechtle, München