

Inhalt

1	Daten und Information – eine Einführung	1
2	Digitale Daten – Bits und Bytes	5
3	Information	9
3.1	Charakteristika des Informationsbegriffes	9
3.2	Die Codierungs-Ebene	10
3.3	Die Syntaktische Ebene der Information	11
3.4	Die Semantische Ebene der Information	14
3.5	Die pragmatische Ebene der Information	15
3.6	Bezüge zwischen den Ebenen	15
3.7	Der Begriff „Information“ in anderen Wissenschaften	17
4	Wissen	17
5	Gründzüge der Shannonschen Informationstheorie	21
6	Diskretisierung (Abtastung) und Quantisierung	28
6.1	Grundsätzliches	28
6.2	Modell der Abtastung	28
7	Datenverarbeitung, Informationsverarbeitung, Wissensverarbeitung	30
8	Gesellschaftliche Wirkungen	32



Daten

Information

Wissen

Ziel dieser Vorlesung ist es einige zentrale Grundbegriffe der Informatik kennen zu lernen. Insbesondere die Begriffe „Daten“, „Information“ und „Wissen“ die zugehörigen Einheiten und Begriffe. Grundlegend für die Informatik ist der Begriff der Abstraktion und die damit einhergehende Modellbildung.

Weitere Zentrale Begriffe sind Digitalisierung, Abtastung und Quantisierung. Die Grundzüge der Shannonschen Informationstheorie sollten verstanden werden.

1 Daten und Information – eine Einführung

Wir beginnen mit einer Diskussion der zentralen Begriffe **Daten, Nachricht und Information**.

Wir werden diese Begriffe im mathematisch/formalen Sinn nicht definieren, hierzu fehlen uns zu viele Grundlagen, aber im „informatischen“ Sinn präzisieren, ihre Verwendung klären, ihre Eigenschaften und ihre Zusammenhänge kennen lernen. Wichtig zu erkennen ist, dass diese Präzisierung sich nicht gänzlich mit dem umgangssprachlichen Gebrauch oder dem Gebrauch in anderen Wissenschaften deckt. Einher damit geht, dass wir viele weitere grundlegende Begriffe kennen lernen werden.

In dieser ersten Einführung können wir keines der Themen vollständig behandeln, wir werden an verschiedenen Stellen auf eine spätere Behandlung in dieser Vorlesung verweisen

müssen oder auch auf andere Veranstaltungen des Informatik-Studiums. Insofern versucht dieses Kapitel auch, erstmals das „Begriffs-Universum“ der Informatik aufzuspannen, die Hauptachsen aufzuzeigen und damit Verbindungen aufzuzeigen, die zwischen den verschiedenen Teilgebieten der Informatik bestehen.

Der Begriff **Daten** (engl. *data*) ist ursprünglich der Plural von Datum (lat. Singular datum, Plural data = Gegebenes). In der Regel wird es auch genau so in der Informatik benutzt.¹ Daten sind (maschinen-) lesbare und bearbeitbare **Repräsentationen von Informationen**: die (abstrakte) Information ist im Rechner (Computer) und in seiner Peripherie durch (konkrete) Daten repräsentiert. Dabei ist die Zuordnung (Relation) zwischen Datum und Information im Allgemeinen nicht eindeutig. Ein und das gleiche Datum kann also verschieden interpretiert werden, also verschiedene Informationen repräsentieren. Entscheidend ist also eine fest vereinbarte Interpretationsvorschrift α welche den abstrahierenden Schritt beschreibt.

$$D \xrightarrow{\alpha} I$$

Dieser Zusammenhang sagt zunächst erst einmal nichts darüber aus, wie D und I konkret aussehen. Für einen (Digital-)Rechner dürfen wir aber annehmen das D eine **Folge** von Binärwerten (0 oder 1, **Bits**), ist. z.B.:

D = 01011101

Die Informationen hingegen sind typischerweise durch eine Folge von **Zeichen** repräsentiert, deren Elemente einer bestimmten Menge entnommen sind, dem **Alphabet**. Alle „zulässigen“ (= gültigen) Informationen (= Zeichenfolgen) gehorchen einer **Syntax** (Regelwerk) und haben eine **Semantik** (Bedeutung). Sie bilden damit eine **Sprache**, zu der eine Pragmatik gehört. Hierzu im Weiteren einiges mehr. Machen Sie sich klar, auch die übliche Dezimalschreibweise von Zahlen wird hierdurch abgedeckt:

I = ‚ANTON‘ (ein Wort über dem Alphabet der lateinischen Großbuchstaben)

I = 42 (ein Wort über dem Alphabet der Ziffern [0,...,9])

¹ Umgangssprachig ist das allerdings anders. Die Bedeutung von "Datum" ist im allgemeinen Sprachgebrauch auf "Kalenderdatum" eingeeengt. Deshalb ist die Verwendung des Wortes "Daten" als Plural von "Datum" unüblich. "Datum" wird behandelt, als ob es ein Singularetantum (Wort ohne Plural) wäre; "Daten" wird behandelt, als ob es ein Pluraletantum (Wort ohne Singular) wäre. Um dagegen das Wort "Datum" als Singular zu "Daten" zu vermeiden, werden Wörter wie "Datenelement", "Angabe" oder "Wert" verwendet.

Im Englischen ist dies anders: „Is 'data' singular or plural? Strictly speaking, data is the plural of datum, and should be used with a plural verb (like facts). However, there has been a growing tendency to use it as an equivalent to the uncountable noun information, followed by a singular verb. This is now regarded as generally acceptable in American use, and in the context of information technology. The traditional usage is still preferable, at least in Britain, but it may soon become a lost cause.“ [aus Ask Oxford: <http://www.askoxford.com/asktheexperts/faq/aboutgrammar/data?view=uk/> , 30.8.2005]

In der Nachrichtentechnik und Übertragungstechnik und auch in der klassischen Informationstheorie benutzt man anstelle **Daten** den Begriff **Nachricht**. Das Verhältnis zur Information ist genau dasselbe:

$$N \xrightarrow{\alpha} I$$

Um übertragbar und/oder speicherbar zu sein, muss eine Nachricht eine Repräsentation haben, in der Digitaltechnik wieder eine Folge von Binärwerten (Bits). Häufig spricht man hier dann von **Signal S**, das eine Nachricht repräsentiert und dieses kann prinzipiell sehr vielgestaltig sein, insbesondere also **digital** oder auch **analog** (kontinuierliche Größe, z.B. eine Spannung aus einem bestimmten Intervall). Wir erhalten also folgenden Zusammenhang:

$$S \xrightarrow{\eta} N \xrightarrow{\alpha} I$$

Daten und Nachrichten sind also prinzipiell beliebig lange Folgen von Bits. Aus organisatorischen oder technischen Gründen werden oft auch größere Einheiten von Bits betrachtet (Nibbels, Oktetts \approx Bytes, Worte, Blöcke, etc). Dieses sind Binärvektoren einer festen Länge. Soll eine Nachricht von einem Sender zu einem Empfänger gelangen, muss zwischen beiden ein **Übertragungskanal** bestehen, z.B. eine Verbindung über ein Rechnernetz.

Die oben angegebene Interpretationshierarchie können wir erweitern: interpretierbare (und anwendbare) Nachrichten sind Informationen; interpretierbare (und anwendbare) Informationen nennen wir auch **Wissen W**. Also

$$S \xrightarrow{\eta} \begin{matrix} N \\ D \end{matrix} \xrightarrow{\alpha} I \xrightarrow{\beta} W$$

Im Zentrum unserer weiteren Diskussion werden zunächst (in PRG) Daten und Informationen stehen und die Interpretationsvorschrift α . Unsere Darstellung ist bisher noch sehr allgemein. Wir wollen einige Beispiele betrachten und daraus entsprechende grundlegende Eigenschaften und Forderungen ableiten. Festzuhalten ist, dass wir mit dem oben dargestellten Zusammenhängen eine ganz charakteristische Arbeitsweise der Informatik in einem ersten Beispiel kennen gelernt haben. Das sind die Prinzipien der Abstraktion und Modellbildung: Die Daten sind dabei das konkrete, das Reale – Information ist das Modell:

$$D \xrightarrow{\alpha} I$$

(reale) Welt	Modell
konkret	abstrakt

In beiden Welten, der realen Welt und der Modellwelt bestehen bestimmte Regeln und interne Zusammenhänge. Die Relation α beschreibt die Abstraktion. **Abstrahieren** heißt immer: **Details** (z.B. der Implementierung, der Realisation ...) **weglassen** um damit in der Regel eine Vereinfachung zu erreichen, die es uns leichter macht, bestimmte Sachverhalte zu verstehen oder zu ändern, zu manipulieren. Nur wenn die Relation α (die Abstraktion) bestimmte Eigenschaften hat, lassen sich aus der Betrachtung des Modells Rückschlüsse auf die reale Welt ziehen.

Gehen wir von der Information als gegeben aus und suchen dafür eine Repräsentation als Daten, so bezeichnen wir diesen Vorgang als Kodierung. Unter **Kodierung** (Codierung) versteht man allgemein die Zuordnung (oder Abbildung) der Werte eines Zeichenvorrats auf Werte eines anderen Zeichenvorrats. Unter Zeichen verstehe man eine Ausprägung (Form, Wert) eines Signals (insbesondere auf der Ebene der Daten) oder auch von Symbolen (auf der Ebene der Information). Der Zeichenvorrat (Alphabet) ist die Menge der zulässigen Zeichen (Werte, Signale, Symbole) die vorkommen können.

Kodierung erfolgt zu einem bestimmten Zweck:

- Speicherung
- Übertragung
- Berechnung
- Komprimierung
- Verschlüsselung
- Veranschaulichung

Betrachtet man die oben dargestellten Zusammenhänge, so wird klar, dass Informationen durch Kodierung zu Daten werden (insbesondere zur Speicherung und Berechnung), zu Nachrichten (insbesondere zur Übertragung, ggf. verschlüsselt) und zu anderen Informationen, z.B. zur Veranschaulichung.

Eine Kodierung ist z.B. notwendig, da die **für den Menschen verständliche Information** auf eine für den Rechner geeignete oder speicherbare, bearbeitbare Darstellung abgebildet werden muss. Hier ist dies eine Zuordnung von Symbolen, die für den Menschen Sinn machen, zu Bitfolgen. Die einer Codierung zugrunde liegende Abbildung muss **berechenbar, eindeutig und (in der Regel) umkehrbar sein**. Dieses werden wir später noch genauer betrachten.

Hierzu betrachten wir im Weiteren zunächst einmal die Welt der Daten, speziell im Digitalrechner, dann die Welt der Informationen, also die Grundlagen der formalen Sprachen und dann die Interpretationsvorschriften und weitere Zusammenhänge.

Aber zunächst einige Beispiele (um den trockenen Formalismus der Informatik) die „Macht“ zu nehmen:

Sprachliche Nachrichten sind zum Beispiel

1. bis morgen / see you tomorrow
2. Ta2-c2 / Desoxyribonukleinsäure
3. Seelöwe gesichtet
4. Lirpa

1. zeigt zwei verschiedene Nachrichten, die normalerweise die gleiche Information übermitteln sollen.
2. Keine Geheimsprache, sondern eine Fachsprache
3. Könnte ein vereinbartes Stichwort zur Auslösung einer Handlung sein: maskierte Geheimnachricht
4. sehr einfache Verschlüsselung: rückwärtsgelesen: April.

2 Digitale Daten – Bits und Bytes

In einem Digitalrechner (genauer müsste dieser ja eigentlich „Binärechner“ heißen, aber das ist vollkommen ungebräuchlich ;-)) werden Daten durch Folgen von **Bits** repräsentiert.

Ein **Bit** lässt zwei mögliche Werte auf eine Frage zu, z.B. Ja oder Nein, Wahr oder Falsch, Links oder Rechts. Oft werden in diesem Zusammenhang die beiden **Werte 0 und 1 benutzt**, früher auch 0 und 1.

Bit ist eine Wortkreuzung aus **binary digit**, englisch für Binärziffer. Der Begriff wurde von dem Mathematiker John W. Tukey vermutlich 1946, nach anderen Quellen schon 1943 vorgeschlagen. In der Fachliteratur wurde der Begriff 1948 zum ersten Mal von Claude Shannons in seiner berühmten Arbeit „A Mathematical Theory of Communication“ erwähnt.

Technisch werden die beiden Werte eines Bits, man sagt in der Regel Zustände, z.B. folgendermaßen realisiert

- Die Stellung eines Kippschalters mit zwei Zuständen, zum Beispiel mit den Stellungen EIN oder AUS.
- Der Schaltzustand eines Transistors, "geringer Widerstand" oder "hoher Widerstand".
- Das Vorhandensein einer Spannung, die größer oder kleiner als ein vorgegebener Wert ist, z.B. $[0..0,8V] = 0$ und $[2,4..5V] = 1$ (Werte im Bereich $(0,8...2,4V)$ sind undefinierte fehlerhafte Zustände.
- Magnetisierung (auf einer Festplatte)
- Loch oder Vertiefung in einer CD (Compact Disk)

Sowohl der Gebrauch als auch die Schreibweise (Bit oder bit) sind dabei uneinheitlich und variieren in der Literatur. Allen Verwendungsweisen im Zusammenhang mit Information und Informationstechnik ist gemeinsam, dass ein Bit oder bit als ein Maß für die Größe bzw. die Menge von Daten oder auch Informationen betrachtet wird.

Mögliche Verwendungen für bit/Bit sind (bitte achten Sie auf die Groß- und Kleinschreibung)

- als Element in einer Folge von Binärwerten, meist **Bit**, z.B. Bit 3
- als Stelle in einem nach dem Stellenwertsystem codierten Binärwort meist **Bit** (von z. B. 32 Bit)
- als Speicherzelle, meist **Bit**
- als Einheit für eine Datenmenge (siehe auch Nit oder Hartley), meist **bit**
- als Einheit für den Informationsgehalt (siehe auch Shannon), meist **bit**
- [als Datentyp in einigen Programmiersprachen für eine 1 Bit breite Einheit, die True (Wahr) oder False (Falsch) sein kann, dann **Bit**]

Achtung: Der Gebrauch der Groß- und Kleinschreibung ist weder in der deutschen noch in der englischen Fachliteratur einheitlich. Das Wort **Bit** wird üblicherweise großgeschrieben, wenn es sich um die Bezeichnung „physikalischer“ Bits handelt. Zum Beispiel: Der Datenbus besitzt eine Breite von 16 Bit. Sonst in der Regel klein, z.B. zur

Angabe von Datenraten (bit pro Sekunde). Ethernet 100BaseT mit einer Datenrate von 100 Mbit/s.²

Größere Werte werden durch Präfixe nach folgender Tabelle angegeben:

Mehrfaches von Bits					
SI Internationales Einheitensystem ISO/IEC			Binär-Einheiten IEC 60027-2, Second edition, 2000-11		
Name	Symbol	Mehrfaches	Name	Symbol	Mehrfaches
kilobit	kbit	10^3	kibibit	Kibit	2^{10}
megabit	Mbit	10^6	mebibit	Mibit	2^{20}
gigabit	Gbit	10^9	gibibit	Gibit	2^{30}
terabit	Tbit	10^{12}	tebibit	Tibit	2^{40}
petabit	Pbit	10^{15}	pebibit	Pibit	2^{50}
exabit	Ebit	10^{18}	exbibit	Eibit	2^{60}
zettabit	Zbit	10^{21}			
yottabit	Ybit	10^{24}			

Anmerkungen: Die angegebenen Präfixe sind die strenge (und korrekte) Version. Man muss jedoch feststellen, dass sich die IEC Norm auch im Fachjargon noch nicht allgemein durchgesetzt hat. Also wird man anstelle von gibibit (Gibit) = 2^{30} bit = $(2^{10})^3$ bit sehr häufig „einfach“ gigabit (Gbit) hören oder lesen, unterstellend, dass der „Nachrichtenempfänger“ auch Fachmann ist und durch den Kontext oder den Wert selbst schon das richtige interpretiert. Dazu noch einige historische Anmerkungen:

„Once upon a time, computer professionals noticed that 2^{10} was very nearly equal to 1000 and started using the SI prefix "kilo" to mean 1024. That worked well enough for a decade or two because everybody who talked kilobytes knew that the term implied 1024 bytes. But, almost overnight a much more numerous "everybody" bought computers, and the trade computer professionals needed to talk to physicists and engineers and even to ordinary people, most of whom know that a kilometer is 1000 meters and a kilogram is 1000 grams.

Then data storage for gigabytes, and even terabytes, became practical, and the storage devices were not constructed on binary trees, which meant that, for many practical purposes, binary arithmetic was less convenient than decimal arithmetic. The result is that today "everybody" **does not** "know" what a megabyte is. When discussing computer memory, most manufacturers use megabyte to mean $2^{20} = 1\,048\,576$ bytes, but the manufacturers of computer storage devices usually use the term to mean 1 000 000 bytes. Some designers of local area networks have used megabit per second to mean 1 048 576 bit/s, but all telecommunications engineers use it to mean 10^6 bit/s. And if two definitions of the megabyte are not enough, a third megabyte of 1

² Um einer weiteren möglichen Verwirrung oder Verwechslung vorzubeugen, sei darauf hingewiesen, dass Bit oder bit unterschieden werden muss vom Qubit (Quantenbit), das in der Quanteninformatik/bei Quantencomputern verwendet wird: Ein **Qubit** (für "Quantenbit") ist ein beliebig manipulierbares Zweizustands-Quantensystem. Als Zweizustands-Quantensystem ist das Qubit das einfachste nichttriviale Quantensystem überhaupt. Zweizustandssystem bedeutet hierbei, dass jede Messung des Systems dieses nur in einem von zwei (durch die gewählte Messgröße bestimmten) Zuständen finden kann. Es bedeutet jedoch *nicht*, dass das System selbst nur zwei verschiedene Zustände annehmen kann. In der Tat kann jedes nichttriviale quantenmechanische System prinzipiell unendlich viele verschiedene Zustände annehmen.

024 000 bytes is the megabyte used to format the familiar 90 mm (3 1/2 inch), "1.44 MB" diskette. The confusion is real, as is the potential for incompatibility in standards and in implemented systems.

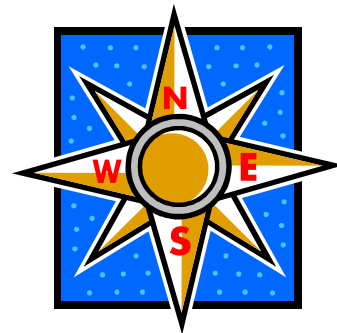
Faced with this reality, the IEEE Standards Board decided that IEEE standards will use the conventional, internationally adopted, definitions of the SI prefixes. Mega will mean 1 000 000, except that the base-two definition may be used (if such usage is explicitly pointed out on a case-by-case basis) until such time that prefixes for binary multiples are adopted by an appropriate standards body." [nach <http://physics.nist.gov/cuu/Units/binary.html> , 30.8.2005].

Zwischenruf: Schauen Sie sich die Übersicht zu nationalen und internationalen Normungsgremien der Informatik und Informationstechnik in der Vorlesung 0 noch einmal an.

Bitfolgen werden zur Darstellung von mehr als zwei Werten benötigt (z.B. die Beantwortung einer Frage enthält mehr als zwei mögliche Antworten, wie "Woher kommt der Wind?").

Daten (Kode) Information

000	→	Nord
001	→	Nordost
010	→	Ost
011	→	Südost
100	→	Süd
101	→	Südwest
110	→	West
111	→	Nordwest



Jedes zusätzliche Bit in einer Bitfolge verdoppelt die Anzahl der möglichen, unterschiedlichen Bitfolgen, d.h. es gibt genau 2^N mögliche Bitfolgen der Länge N.

Moderne Computer und Speichermedien verfügen über Speicherkapazitäten von Milliarden von Bits. Speichergrößen werden daher in anderen Einheiten angegeben. Im Allgemeinen verwendet man hier das Byte (ein Oktett von acht Bit) als Grundeinheit und Potenzen von 2^{10} (= 1024) als Einheitenpräfixe.

Im Bereich der Datenübertragung hat sich jedoch das Bit als Grundeinheit bei der Angabe der Datenübertragungsrate gehalten - ISDN überträgt 64 kbit/s (64.000 Bit pro Sekunde) auf einem Kanal, Fast Ethernet 100 Mbit/s (100 Millionen Bit pro Sekunde). Anders als beim Byte hält man sich hier streng an das SI-System für Vorsilben.

Daneben wird das Bit als Einheit verwendet:

- für die Angabe der Kapazität *einzelner* Speicherchips (hier allerdings mit Binärpräfixen); Beispiel: ein 512-Mb-Chip (Megabit, nicht zu verwechseln mit MB Megabyte) speichert 2^{29} Bits = 2^{26} Bytes, also 64 MB, davon acht Stück auf einem Speicherriegel ergibt ein 512-MB-DIMM Als (**D**ual **I**n-**L**ine **M**emory **M**odule) werden Speichermodule für den Arbeitsspeicher von Computern bezeichnet. Im Gegensatz zu Single Inline Memory Modulen (SIMM) haben

DIMMs zwei getrennte Reihen von Kontaktpins - eine auf der Vorderseite der Leiterplatte und eine auf der Rückseite.

- für Busbreiten bzw. die Verarbeitungsbreiten (Maschinenworte) in Rechensystemen auf Chip oder Systemebene.

Bit ist nicht die einzige übliche Einheit für die Datenmengen und Speicherzellen. Gebräuchlich sind insbesondere Byte und ggf. noch größere Einheiten wie (Maschinen-)Worte, Blöcke, Pakete, etc..

Ein **Byte** ist eine Bezeichnung für:

- (fach-)umgangssprachlich die eines Tupels von 8 Bit, deren formale ISO-konforme Bezeichnung Oktett wäre. (1 Byte = 8 Bit)
- eine adressierbare Speichereinheit in einem Rechensystem
- einen Datentyp in einigen Programmiersprachen für eine 8 Bit breite Einheit, die 256 mögliche Zustände annehmen kann
- eine Maßeinheit für 8 Bit bei Größenangaben für Datenmengen (Kapazitäten) (1 Byte = 2^3 Bit = 8 Bit) (nach: IEC 60027-2, Second Edition, 2000-11).
- eine Datenmenge von zwei *Nibbles*, die 256 verschiedene Werte annehmen kann und durch zwei hexadezimale Ziffern (00 bis FF) darstellbar ist (1 Nibble = 4 Bit = 1/2 Byte; 2 Nibbles = 8 Bit = 1 Byte)

Bei den meisten heutigen Rechensystemen fallen alle diese Definitionen in eins zusammen.

Das Wort Byte ist künstlich und stammt von englisch bit (deutsch: bisschen) **und** bite (deutsch: Happen). Verwendet wurde es, um eine Speichermenge oder Datenmenge zu kennzeichnen, die ausreicht, um ein alphanumerisches Zeichen (Buchstaben, Zahlen, Sonderzeichen) darzustellen. Der Begriff wurde 1956 von Werner Buchholz geprägt in einer frühen Designphase eines IBM-Computers. Im Original beschrieb Bite eine Breite von 6 Bit und stellte die kleinste direkt adressierbare Speichereinheit dieses Computers dar. Bereits Ende 1956 erfolgte aber der Übergang zu 8 Bit. Die Schreibweise Bite wurde zu Byte geändert, um versehentliche Verwechslungen mit Bit zu vermeiden.

Byte wird auch als Abkürzung für „binary term“ und Bit als Abkürzung für „binary digit“ erklärt – dabei handelt es sich aber um Backronyme³. Eine weitere Erklärung des Begriffes liegt in der Kurzform "**by** eight", frei übersetzt "achtfach".

³ Als **Backronyme** werden Akronyme oder Abkürzungen bezeichnet, die nachträglich einen neuen Bedeutungsinhalt bekommen haben. *Backronym* ist dabei eine Wortkreuzung aus „rückwärts“ (englisch: *back*), also erst nach der Abkürzung selbst entstanden, und Akronym (englisch: *acronym*).

Ein humorvolles Beispiel für ein Backronym ist die Interpretation von AEG als „**A**uspacken – **E**inschalten – **G**eht nicht“ (eigentlich: Allgemeine Elektrizitäts-Gesellschaft). Der Hersteller versuchte in der Werbung zu kontern „**A**us **E**rfahrung **G**ut“. ;-)

Beispiele aus der Informatik sind:

- **DVD** – „Digital Versatile Disc“ (*digitale, vielseitige Scheibe*) - ursprünglich „Digital Video Disc“ (*digitale Videoscheibe*)
- **ISDN** – „Integrated Services Digital Network“ – ursprünglich „Integriertes Sprach- und Datennetz“
- Commodore **PET** – *Personal Electronic Transactor* bezeichnet – ursprünglich ist nicht bekannt Das englische Wort *pet* bedeutet auch "Haustier" oder "Liebling". Böse Zungen behaupten, dass der Name auch für **P**eddles (der Entwickler) **E**go **T**rip stehen könnte.

Viele Abkürzungen werden leider auch verwendet, ohne deren eigentliche Bedeutung genau zu kennen, wodurch sich deren falscher Gebrauch durchsetzt. So stößt man heutzutage auch in renommierten

Zur Unterscheidung der ursprünglichen Bedeutung als kleinste adressierbare Einheit und der Bedeutung als 8-Bit-Tupel wird in der Normungs-Fachliteratur (abhängig vom Fachgebiet) korrekterweise auch der Begriff **Oktett** für letzteres benutzt, um eine klare Trennung zu erzielen.

Für die üblichen Präfixe gilt das gleiche, wie schon für bit ausgeführt:

Der Unterschied zwischen GByte und GiByte hat wohl tatsächlich bei vielen Nutzern (insbesondere den „Everybodies“) Verwunderung oder gar Ärger ausgelöst: Man kauft eine 80 GByte Platte und eingebaut im Computer zeigt das Windows XP Betriebssystem: Speicherkapazität 74,5 GB (eine weitere Schreibweise!). Wer hat da möglicherweise getäuscht oder betrogen, das sind immerhin 5,5 GB(yte) oder fast 7% weniger als gekauft? ;-)

3 Information

Information (von lateinisch: *informare* 'bilden, durch Unterweisung Gestalt geben') allgemein ist ein potenziell oder tatsächlich vorhandenes nutzbares oder genutztes Muster von Materie und/oder Energieformen, die für einen Betrachter innerhalb eines bestimmten Kontextes relevant ist. Wesentlich für die Information ist die **Wiedererkennbarkeit** sowie der **Neuigkeitsgehalt**. Das verwendete Muster verändert den Zustand eines Betrachters – im menschlichen Zusammenhang insbesondere dessen **Wissen**.

Formaler ist **Information** die **Beseitigung von Unbestimmtheit**.

3.1 Charakteristika des Informationsbegriffes

Information ist heute ein sehr weitläufig verwendeter und daher auch sehr schwer abzugrenzender Begriff. Verschiedene Wissenschaften betrachten die Information als ihr Arbeitsgebiet, namentlich die Informatik, die Informationstheorie, die Nachrichtentechnik und auch die Semiotik (allgemeine Lehre von den Zeichen, Zeichensystemen). Dabei sind die Begriffe meist dieselben oder ähnlich, aber die Bedeutungen je nach Wissenschaft durchaus unterschiedlich!

Im allgemeinen Sprachgebrauch sowie in einigen Wissenschaften (Semiotik, Informationswissenschaften) wird "Information" mit "Bedeutung" oder "übertragenem Wissen" gleichgesetzt.

Information ist typischerweise durch eine Folge von **Zeichen** repräsentiert, deren Elemente einer bestimmten Menge entnommen sind, dem **Zeichenvorrat**. Beispiele für Zeichenvorräte sind:

- Kleinbuchstaben: {a, b, c, d, e, f, g, ... ,z}

Fachzeitschriften auf das Wort LCD-Display, obwohl das "D" in der Abkürzung bereits für "Display" steht (Liquid Crystal Display). Ähnlich verhält es sich mit einer für Grafikkarten vorgesehene PC-Schnittstelle (Accelerated Graphics Port), welche oft fälschlicherweise als AGP-Port bezeichnet wird. Diese Verdoppelungen sind streng genommen Tautologien (also weiße Schimmel)

- Desoxyribonukleinsäuren (DNS) {A, C, G, T}
Adenin, Cytosin, Guanin, Thymin
- Boolesche Werte {0, 1} {false, true}

Wenn auf dem Zeichenvorrat eine **Ordnung** definiert ist, das heißt eine fest definierte Reihenfolge der Zeichen, so sprechen wir von einem **Alphabet**, wie zum Beispiel bei den Buchstaben. In einem Alphabet lässt sich für jeweils zwei Zeichen ein Vergleich durchführen, z.B. 'a' ist kleiner als 'b' oder 'c' ist größer als 'b'.

Alle „zulässigen“ (= gültigen) Informationen (= Zeichenfolgen/Wörter) gehorchen einer **Syntax** (Regelwerk) und haben eine **Semantik** (Bedeutung). Sie bilden damit eine **Sprache**, zu der eine Pragmatik gehört. Hierzu im Weiteren einiges mehr.

Basiselemente sind also Zeichen, aus denen gemäß einer Syntax durch Konkatenation Wörter entstehen die eine Semantik tragen. Besonders im Kontext natürlicher Sprachen ist der Sprachgebrauch etwas anders: Hier werden aus den elementaren Wörtern Sätze gebildet, die eine Semantik tragen und somit eine Sprache bilden.

Es lassen sich also vier Ebenen erkennen:

- **Codierung**
- **Syntax**
- **Semantik**
- **Pragmatik**

Diese Ebenen steigern sich im Hinblick auf den Bedeutungsgehalt der Information. Die Codierungs-Ebene behandelt die Abbildung von Informationen auf Daten oder Nachrichten; die Syntaxebene repräsentiert die Sichtweise der Linguistik und/oder die der Theorie der formalen Sprachen, die semantische Ebene integriert Ansätze aus der Semiotik und die Pragmatik greift auf Konzepte der Kognitionswissenschaften zurück.

3.2 Die Codierungs-Ebene

Haben wir einleitend schon besprochen:

$$D \xrightarrow{\alpha} I \text{ oder } N \xrightarrow{\alpha} I,$$

d.h. jedem Datum oder jeder Nachricht wird durch eine Abbildungsvorschrift (gegeben z.B. in einer Kodierungstabelle) eine Information zugeordnet.

Hier setzt die **Informationstheorie** an. Diese wurde 1948 von **Claude Shannon** in seinen Grundzügen vorgestellt. Er betrachtet insbesondere statistische Aspekte der Zeichen und der Codes. (Die Bedeutung der Information geht bei Shannon nur implizit in den Wahrscheinlichkeiten der verwendeten Zeichen ein.) Das unmittelbare Ziel seiner Überlegungen war die optimale Übertragung von Information in einem Nachrichtenkanal (z.B. Telefonie, Funk). Grundlegende Fragen für ihn waren:

- Wie kodiere ich Informationen optimal (d.h. z.B. mit minimaler Datenmenge) oder allgemeiner: Fragen nach der Wahl von optimalen Codes für einen bestimmten Zweck
→ **Kodierungstheorie**

- Wie groß ist die Übertragungskapazität eines Kanals, d.h. wie viele Daten resp. Information kann ich maximal pro Zeiteinheit übertragen, z.B. über eine Telefonleitung
→ **Abtasttheorie**
- und damit mittelbar auch: Was ist Information (in diesem statistischen Kontext)?

Die Antworten sind grundlegend für die gesamte Informatik und wir werden uns unten noch detaillierter damit befassen.

3.3 Die Syntaktische Ebene der Information

Auf der syntaktischen Ebene wird **Information nur als Struktur gesehen**. Die Bedeutung der Information ist hierbei weitgehend uninteressant. Insbesondere sind dies Regeln, wie aus Zeichen Worte zu bilden sind: Das nennen wir eine **Grammatik oder Syntax**.

Grundprinzip der syntaktischen Information ist die **Unterscheidbarkeit**: Information enthält, was unterschieden werden kann. Eine Unterscheidung setzt jedoch mindestens zwei unterschiedliche Möglichkeiten voraus.

Eine **Syntax** (griechisch σύνταξις [syn'taksis] - die Zusammenstellung) behandelt die Muster und Regeln, nach denen Wörter (respektive Zeichen) zu größeren funktionellen Einheiten wie Phrasen (Teilsätze) und Sätzen (respektive Wörtern) zusammengestellt und Beziehungen wie Teil-Ganzes, Abhängigkeit etc. zwischen diesen formuliert werden (**Satzbau**). In der allgemeinen Linguistik also Wörter → Sätzen, in der Informatik meist Zeichen → Wörtern.

In der Informatik wird eine formale Grammatik benutzt, um bei einer formalen Sprache (z. B. Programmiersprache) erlaubte Konstruktionen festzustellen und unerlaubte auszuschließen. Die syntaktische Definition einer Sprache wird meist in einer formalen Notation heute meist in einer Erweiterten Backus-Naur-Form angegeben.

Die Backus-Naur-Form oder Backus-Normalform, kurz BNF, ist eine kompakte formale Metasyntax (Metasprache), die benutzt wird, um kontextfreie Grammatiken darzustellen. Formale Grammatiken sind mathematische Modelle von Grammatiken, die so genannte formale Sprachen erzeugen. Hierzu zählt insbesondere die Syntax höherer Programmiersprachen. Sie kann aber auch zur Beschreibung der Syntax von Befehlssätzen, Kommunikationsprotokollen oder allgemein Information verwendet werden.

Ursprünglich war sie nach John Backus⁴ benannt, später wurde sie (auf Anregung von Donald E. Knuth⁵) auch nach Peter Naur⁶ benannt. Beide waren Informatikpioniere, die

⁴ John Warner Backus (1924 -) ist einer der Pioniere der Informatik und vor allem durch die nach ihm benannte Backus-Naur-Form bekannt. 1949 entwickelte er für IBM die Programmiersprachen Speedcoding und 1954 Fortran. Aus der Beschäftigung mit Grammatiktheorie entstand die Backus-Naur-Form. 1977 erhielt er dafür den Turing-Preis. In seinem bekannten Paper *Can Programming Be Liberated from the von Neumann Style? A Functional Style and Its Algebra of Programs* empfahl Backus die Hinwendung zum Paradigma der funktionalen Programmierung. Mit FP entwickelte er eine der ersten funktionalen Programmiersprachen.

⁵ Donald Ervin Knuth (1938 -) ist einer der sehr bekannten US-Informatiker (Stanford). Er verfasste (und verfasst noch) *The Art of Computer Programming* und schuf eigens dafür mit **TeX und Metafont** Computerprogramme, welche druckreifen Textsatz ermöglichen und besonders im mathematisch-akademischen Bereich ihren Einsatz finden. 1974 erhielt er den Turing-Preis

⁶ Peter Naur (1928 -) ist dänischer Informatik-Pionier. Er selbst mag ungern mit der Backus-Naur-Form in Verbindung gebracht werden (dies geschah durch Donald Knuth) und er sagt, dass er bevorzugen würde, sie *Backus-Normal-Form* zu nennen. 2005 erhielt er den Turing-Preis

sich mit der Erstellung der Algol-60-Regeln und insbesondere mit dem Compilerbau beschäftigten. Durch die Backus-Naur-Form im Algol 60 Report wurde es erstmals möglich, die Syntax einer Programmiersprache formal exakt, also ohne die Ungenauigkeiten natürlicher Sprachen, darzustellen.

Eine formale Grammatik⁷ erlaubt es zu entscheiden, ob ein Text der Grammatik folgt, also ob er „gültig“ ist. Die Menge aller Texte, die der Grammatik folgen, nennt man die **Sprache** dieser Grammatik. Einen Text, der der Grammatik folgt, nennt man auch ein **Wort** dieser Sprache. Umgekehrt erlaubt es eine Grammatik auch, alle Wörter ihrer Sprache ausgehend von dem Startsymbol zu *erzeugen*.

Eine formale Grammatik besteht aus

- Terminalsymbolen (*terminals*),
- Nichtterminalsymbolen (*nonterminals*) und
- Ableitungsregeln (Produktionen)
- einem Startsymbol

Ein **Terminalsymbol** bzw. **Terminalzeichen** ist ein Symbol, das (einzeln) nicht weiter durch eine Produktion ersetzt werden kann. Mit anderen Worten, ein Terminalsymbol kommt auf keiner linken Regelseite einzeln vor, ist elementar (siehe nachfolgend). In der Theorie (der Informatik) werden Terminalsymbole häufig mit Kleinbuchstaben

⁷ Formal kann man eine Grammatik so definieren: Eine formale Grammatik $G = (N, \Sigma, P, S)$ ist ein 4-Tupel bestehend aus

- einer endlichen Menge N von **Nichtterminalsymbolen** (auch Nichtterminale oder Variablen genannt)
- dem Alphabet Σ : einer endlichen Menge Σ von **Terminalsymbolen** (auch Terminale)
- einer endlichen Menge P von **Produktionsregeln** (auch Produktionen genannt)
- einem Startsymbol S , das zur Menge der Nichtterminalen gehört ($S \in N$)

Keine der Mengen N , Σ und P ist die leere Menge. Die Menge der Nichtterminalsymbole und die Menge der Terminalsymbole sind disjunkt. (Es gibt also kein Symbol, das sowohl Nichtterminal- als auch Terminalsymbol ist.)

Eine Produktionsregel P besteht aus einer linken (Prämisse) und einer rechten Seite (Konklusion), die jeweils ein Wort (= eine Konkatenation, Aneinanderreihung) bestehend aus Terminalen und Nichtterminalen sind. Die linke Seite muss mindestens ein Nichtterminal beinhalten und die rechte Seite kann dabei im Gegensatz zur linken Seite auch das leere Wort (oft mit ϵ symbolisiert) sein, also das Wort der Länge Null:

$$P \subseteq (N \cup \Sigma)^+ \times (N \cup \Sigma)^* \quad \text{notiert durch } a \leftarrow b \text{ oder } a ::= b \text{ mit } a \in (N \cup \Sigma)^+ \text{ und } b \in (N \cup \Sigma)^*$$

Der Operator $*$ (der Kleene-Stern) bezeichnet die Kleenesche Hülle von $N \cup \Sigma$, das ist die Menge aller durch Konkatenation (Aneinanderreihung) von Elementen aus N und Σ erzeugbaren Termen einschließlich der leeren Menge. Der Operator $^+$ bezeichnet dasselbe ohne die leere Menge.

Eine Regel kann auf ein Wort angewendet werden, wobei ein beliebiges Vorkommen der linken Seite der Regel im Wort durch die rechte Seite der Regel ersetzt wird. Die Eigenschaften von Grammatiken werden in der Theorie der formalen Sprachen (im 4. Fachsemester) genauer untersucht.

gekennzeichnet, z.B. *aaaaabc*; in der Praxis jedoch auch andere Symbole wie Satzzeichen und Schlüsselwörter von Programmiersprachen, z.B. FOR, IF, PROGRAM, usw.

Ein **Nichtterminalsymbol** wird in BNF durch Ableitungsregeln definiert. Nichtterminalsymbole werden mit spitzen Klammern *<...>* umschlossen.

In einer **Ableitungsregel** (Ersetzungsregel) wird die Zeichenfolge *::=* zur Definition benutzt; das Zeichen *|* (vertikaler Strich) zur Kennzeichnung von Alternativen; Folgen werden durch Hintereinanderschreiben von Terminalsymbolen und Nichtterminalsymbolen angegeben

Beispiel: **Alternative:**

<Ziffer außer Null> ::= 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

Eine Ziffer außer Null ist also entweder eine 1 oder eine 2 ... bis 9. Es lassen sich auch Terminalfolgen definieren, also eine Sequenz. Als Elemente dürfen Terminalsymbole und Nichtterminalsymbole auftreten:

Beispiel: **Sequenz:**

<Ziffer> ::= 0 | <Ziffer außer Null>

<Zweistellige Zahl> ::= <Ziffer außer Null> <Ziffer>

Eine Ziffer ist also eine 0 oder eine Ziffer außer Null. Eine zweistellige Zahl ist eine Ziffer außer Null gefolgt von einer Ziffer.

Wiederholungen muss man in BNF über eine Rekursion definieren. Eine Ableitungsregel kann dazu auf der rechten Seite das Symbol auf der linken Seite enthalten.

Beispiel: **Rekursion**

<Ziffernfolge> ::= <Ziffer> | <Ziffer> <Ziffernfolge>

Also: *Eine Ziffernfolge ist eine Ziffer oder eine Ziffer gefolgt von einer Ziffernfolge.*

Soweit die klassische Notation. Es hat sich aber gezeigt, dass die BNF teilweise umständliche Konstrukte benötigt, um z.B. optionale Elemente, also Elemente, die ausgelassen werden dürfen, sowie sich wiederholende Elemente darzustellen. Niklaus Wirth hat in der Definition der Sprache Pascal zusätzliche Elemente eingeführt um diese Schwierigkeiten zu umgehen und nannte dies extended BNF (EBNF, erweiterte BNF); In der Literatur sind verschiedene Varianten der Backus-Naur-Form gebräuchlich, wir benutzen im Weiteren die für **Python** **genutzte modifizierte Form** oder die genormte Form, siehe ISO/IEC 14977: (1996E).

Verwendung	Zeichen	Kommentar
Definition	<i>::=</i>	
Endezeichen	<i><new line></i>	jede Definition steht in einer Zeile; Zeilen die mit <i> </i> beginnen sind Fortsetzungszeilen; Leerzeichen sind normalerweise ohne

		Bedeutung, es sei denn in Anführungszeichen " "
Gruppierung	(...)	
Alternative		
Option	[...]	kein oder einmaliges Auftreten des Terms ...
Optionale Wiederholung	(...)*	kein, ein oder mehrmalige Wiederholung des Terms (...)
Wiederholung	(...)+	ein oder mehrmalige Wiederholung des Terms (...)
Terminalzeichen	"a"	a kann ein einzelnes oder mehrere druckbare ASCII Zeichen sein
Alternative von Terminalzeichen	"1" ... "9"	zwei Terminalzeichen, die durch ... verbunden sind bedeuten: ein beliebiges Zeichen in den angegebenen (ASCII-) Grenzen, inklusive der Grenzen selbst, hier die Ziffern 1, 2,3,4,5,6,7,8 oder 9
Kommentar	(<...>)	

Zur Kennzeichnung einer Syntax benutzen wir im weiteren Skript den Schrifttyp Courier und unterlegen die Zeilen zusätzlich hellblau.

Ein Beispiel:

```
name ::= lc_letter (lc_letter | "_")*
lc_letter ::= "a"... "z"
```

Ein name ist also ein lc_letter gefolgt von einer Sequenz von keinem oder mehreren lc_letter oder einem Unterstrich _. Ein lc_letter ist ein ASCII-Kleinbuchstabe a, b, c, ...,z (kein Umlaut!)

3.4 Die Semantische Ebene der Information

Strukturierte, syntaktische Informationen werden erst verwertbar, indem sie interpretiert werden. Das heißt, zur Strukturebene muss die **Bedeutungsebene** hinzukommen. Dazu muss ein bestimmtes Bezugssystem angelegt werden, um die Strukturen in eine Bedeutung überführen zu können. Im obigen Beispiel muss man also "wissen", was *spannend* bedeutet.

Jedoch ist die Überführung von Syntax in Semantik selten so direkt; in der Regel wird die Information über sehr viele unterschiedliche Zwischen-Codes zu immer höheren semantischen Ebenen verarbeitet: Dabei wird auf den unterschiedlichen semantischen Ebenen wiederum Informationsverarbeitung auf strukturell-syntaktischer Ebene geleistet:

Strukturinformation wird in einem Dekodierungsprozess (= hier Abstraktion) in Semantik (Bedeutung) überführt. Dabei wird Strukturinformation stufenweise über Codes in andere Strukturinformation überführt, wobei sich auf den unterschiedlichen semantischen Stufen jeweils Bedeutung für das verarbeitende System entwickelt.

3.5 Die pragmatische Ebene der Information

Diese Ebene kommt dem umgangssprachlichen Informationsbegriff am nächsten. Die Aussage, dass es spannend ist (die wir nun semantisch richtig interpretiert haben; wir wissen, was diese Botschaft uns sagen will), hat echten Informationscharakter, wenn wir uns mittags um zwölf nach einer durchzechten Nacht noch halb schlaftrunken überlegen, ob wir wirklich in die Vorlesung gehen sollen. Der pragmatische Informationsgehalt der - semantisch exakt gleichen - Aussage ist aber gleich null, wenn wir bereits im Hörsaal sitzen und einer spannenden Erläuterung lauschen. Diese Information bietet uns nichts neues.

Smalltalk ist eine Art des Informationsaustausches, bei dem die offensichtlich über die Sprache ausgetauschten semantischen Informationen so gut wie keine pragmatische Information darstellen - wichtig sind hier die Körpersignale, deren Semantik (Freundlichkeit, Abneigung) wir erkennen und pragmatisch (mag er/sie mich?) verwerten können.

In diesem pragmatischen Sinne ist wesentliches Kriterium von Information, dass sie das Subjekt, das die Information aufnimmt, *verändert*, was konkret bedeutet, dass sich die Information, die potentiell dem Subjekt entnommen werden kann, verändert..

3.6 Bezüge zwischen den Ebenen

Wenn man das Phänomen *Information* betrachtet, sind die vier Ebenen im Zusammenhang zu betrachten. Damit Information stattfindet, sind Vereinbarungen auf allen vier Ebenen notwendig.

Das Verständnis der syntaktischen Ebene war lange Zeit gekennzeichnet durch das **Sender-Empfänger-Modell**: Ein Sender will eine Information dem Empfänger mitteilen. Dazu codiert er seine Information nach bestimmten Prinzipien (beispielsweise als Abfolge von Nullen und Einsen nach dem oben erwähnten Prinzip) in einen Informationsträger, der Empfänger wertet diesen Informationsträger aus, denn auch er kennt den Code, und erhält dadurch die Information (siehe auch: Kommunikation).

Nicht immer ist jedoch ein menschlicher Sender vorhanden, der uns etwas mitteilen will. Ein typisches Beispiel ist die **Messung**: Dem physikalischen System ist es, bildlich gesprochen, völlig egal, was wir von ihm denken. Das Ziel der Messung ist eine Informationsübertragung vom gemessenen System zu dem, der die Messung durchführt (man misst, um etwas über das gemessene System zu erfahren).

Ein Beispiel ist die Geschwindigkeitsmessung per Radarfalle: Das Auto hat keine Intention, seine Geschwindigkeit zu verraten (und der Autofahrer meist auch nicht). Dennoch gewinnt der Polizist durch die Messung Information über die Geschwindigkeit. Die unmittelbare Erzeugung von Information wird damit an einen Apparat *delegiert*.

Grundlegend ist, dass Information, die **an Materie als Informationsträger** gebunden ist, auf **Elektromagnetische Wellen** übertragen werden kann. Diese Information kann, da masselos, dann im Prinzip mit Lichtgeschwindigkeit transportiert werden. Schließlich kann die Information wieder zurück an Materiestrukturen gebunden werden. Ein Beispiel für so einen Übertragungsprozess ist das Telefax. Dabei wird die Information eines bestimmten Schriftstückes mit (fast) Lichtgeschwindigkeit über große Entfernungen transportiert und am Ziel auf ein zweites Schriftstück mit (fast) exakt demselben Informationsinhalt übertragen. Allgemeiner: Um Informationen zu transportieren ist ein Informationsträger nötig.

Zusammengefasst:

- Information führt zu einem Gewinn an Wissen.
- Information ermöglicht die Verringerung von Ungewissheit.
- Information ist übertragbar; in Form von Daten bzw. Signalen
- Information ist ein Ereignis, dass den Zustand des Empfängers bzw. Systems verändert.

Auf Programmiersprachen angewendet bedeutet dies: Programmiersprachen sind künstlich geschaffene Sprachen (formale Sprachen) zur Erstellung eines Programms (Anweisungen für den Computer): Sie repräsentieren insbesondere die Information wie eine bestimmte Datenverarbeitung erfolgen soll und welche Daten verarbeitet werden sollen.

Die Bedeutung eines speziellen Symbols (ein Zeichen (z.B. #) oder eine festgelegte Folge von Zeichen, z.B. das Wort *while*) in einer Programmiersprache nennt man dessen Semantik.

Syntax und Semantik kann man der **Spezifikation**, teilweise auch der Dokumentation der Programmiersprache entnehmen. Eine vollständige semantische Spezifikation einer Programmiersprache ist tatsächlich sehr schwierig und dieses in einem einzigen Kalkül zu fassen ist z.B. heute noch Forschungsgegenstand.

Die Intention des Programmierers über die Gesamtheit des Computerprogramms nennt man Pragmatik.

3.7 Der Begriff „Information“ in anderen Wissenschaften

Zum Abschluss sollen hier die einzelnen Fach- und Forschungsrichtungen zu Wort kommen, die je ihr eigenes Verständnis der Information haben. Deutlich wird dabei der jeweilige Ansatz auf den unterschiedlichen, oben geschilderten Ebenen zwischen der reinen Syntax bis zur Pragmatik, teilweise auch mit der besonderen Betonung des Transportcharakters von Information.

Die **Semiotik** (allgemeine Lehre von den Zeichen, Zeichensystemen und Zeichenprozessen) versteht unter Informationen zweckorientierte Daten, die das Wissen erweitern.

Die **Informationswissenschaft** verwendet den Begriff der Information ähnlich zum semiotischen Ansatz. Für sie sind die Begriffe Wissen und Information von zentraler Bedeutung. Information ist dabei Wissenstransfer beziehungsweise "Wissen in Aktion". Information *entsteht* in diesem Sinne immer nur punktuell, wenn ein Mensch zur Problemlösung Wissen (eine bestimmte Wissenseinheit) benötigt. Diese Wissenseinheit geht als Information aus einem Wissensvorrat in einen anderen über, beispielsweise aus einer Datenbank in den Wissensvorrat eines Menschen. Wissen wird intern *repräsentiert*, Information wird - zum besseren Verständnis für den Informationssuchenden - *präsentiert*.

Information kann als wirtschaftliches Gut angesehen werden, da Information im Unternehmen durch Einsatz anderer Produktionsfaktoren (Menschen, Computer, Software, Kommunikation, etc.) produziert, oder von außen angekauft werden kann. Information hat somit einen Wert, der handelbar ist. Der Wert ergibt sich aus dem Nutzen der Information und den Kosten zur Produktion, Bereitstellung und Weiterleitung. Problematisch hierbei ist, dass der potenzielle Käufer den Wert der Information nicht immer im voraus kennt und sie teilweise erst nachdem er sie erworben hat, bewerten kann. Weiterhin kann man Information auch als Produktionsfaktor verstehen. Information wird somit nicht nur konsumtiv genutzt, sondern kann auch produktiv verwendet werden.

Interessant, aber auch klar: Information kann weitergegeben werden, geteilt oder verteilt werden, ohne dadurch weniger zu werden. Information wird durch Weitergabe also „quasi verdoppelt“. Mit Materie oder Energie geht das nicht.

4 Wissen

Beginnen wir mit ein paar Zitaten:

- *"Wir ertrinken in Informationen und hungern nach Wissen." - John Naisbitt*
- *"Das wichtigste Wissen ist zu wissen, was wichtig ist." - Andreas Tenzer*
- *"Etwas wirklich zu wissen heißt, seine Gründe zu kennen." - Francis Bacon*
- *"Der Glaube ist nicht der Aufgang, sondern das Ende allen Wissens. - Johann Wolfgang von Goethe*

Seit der Antike wird der Begriff „Wissen“ diskutiert. Es ist nicht verwunderlich, dass in verschiedenen Disziplinen durchaus verschiedene Ansichten existieren: Zu nennen sind beispielhaft: Philosophie (insbesondere in der Erkenntnistheorie), (Kognitions-) Psychologie, Betriebswirtschaftslehre und last-not-least Informatik.

Im Deutschen ist „Wissen“ abgeleitet vom althochdeutschen Verb *wissan*, das seinerseits von der Form „gesehen haben“ kommt, was auch etymologisch auf die Begründetheit verweist. Der Begriff Wissen ist durchweg positiv belegt. Begriffe ähnlicher Bedeutung, wie Bildung, Einsicht, Erkenntnis, Know-how, Überzeugung, Verstehen, Weisheit, usw. zeigen dies. Gegenworte sind eher negativ: Halbwissen, Pseudowissen, Scheinwissen, Unwissen.

Das was wir betreiben und Sie lernen wollen, Wissenschaft (= Wissen hervorbringende (Wissen schaffende), forschende (Geistes-)Tätigkeit), ist also gesellschaftlich eine sehr hoch bewertete Tätigkeit.

Wissen steht in der griechischen Philosophie im Gegensatz zur Meinung. Demnach impliziere Wissen Wahrheit und könne durch keine Argumentation widerlegt werden, während eine Meinung zwar wahr sein könne, aber diskutabel sei. Diese Ansicht gilt auch in der Philosophie aber nicht unumstritten. Wir folgen aber weitgehend diesem

Wissen unterscheidet sich vom *bloßen* Glauben oder *bloßer* Meinung:

- **Wissen** = subjektiv und objektiv zureichendes Fürwahrhalten
- **Glauben** = subjektiv zureichendes, objektiv unzureichendes Fürwahrhalten
- **Meinen** = subjektiv und objektiv unzureichendes Fürwahrhalten

Als eine Art kleinster gemeinsamer Nenner lassen sich **drei Aussagen** formulieren, die **alle drei erfüllt** sein müssen!

- Dem Wissen liegen Informationen zugrunde, oder anders formuliert: Wissen ist mit Bedeutung (Semantik) und Pragmatik ausgestattete Information.
- Diese Informationen müssen derart aufeinander bezogen sein, dass sie nachvollziehbar in sich stimmig sind, also ohne Widersprüche!
- Neben der inneren Übereinstimmung muss sich Wissen in Übereinstimmung mit den wahrnehmbaren Bedingungen einer Umwelt (der realen Welt, aber auch der sozialen Gemeinschaft: Wissen ist kein individuelles Relikt) befinden.

Alle genannten Punkte sind diskutabel. Trotzdem treffen sie einen Kern: **Wissen ist mehr als nur Information**. Mir gefallen folgende Zusammenfassungen sehr:

- **Informationen sind anwendbare/nutzbare Daten**
- **Wissen ist anwendbare/nutzbare Information**

Nach unseren Definitionen unterscheiden sich Daten und Informationen durch die zugehörige Interpretationsvorschrift, die wir kennen müssen, damit die Daten nutzbringend verwendet werden können. Offen ist, ob nach den obigen Feststellungen auch zwischen Informationen und Wissen eine derartige einfache Relation besteht! Die oben formulierten Aussagen gehen deutlich darüber hinaus!

Zentraler Punkt der Informatik-Aktivitäten ist die Frage, **wie Wissen im Computer repräsentiert werden kann**. Hierzu ist es sinnvoll, sich mit den verschiedenen Dimensionen des Wissens auseinanderzusetzen: Wir unterscheiden:

- deklaratives Wissen versus prozedurales Wissen
- diskursives Wissen versus narratives Wissen
- explizites Wissen versus implizites Wissen

- informal – semi-formal – formal
- usw.

Vor allem in der **Kognitionspsychologie** unterscheidet man das **deklaratives Wissen** („Wissen *was*“) in Abgrenzung zu **prozeduralem Wissen** (handlungsorientiertem Wissen („Wissen *wie*“)).

Das deklarative Wissen hat wiederum viele Unterformen: Wissen über Fakten, Konzepte, semantische Beziehungen, Ereignisse und Wirkungen, einschränkende Bedingungen, Metawissen, usw.

Für die Informatik sehr wichtig ist die Unterscheidung in implizites Wissen und explizites Wissen und die damit einhergehenden Umwandlungsprozesse (nach Nonaka & Takeuchi)

Implizites Wissen (tacit knowledge) ist schwer kommunizierbar, kaum formalisierbar, in Köpfen von Personen gespeichert (embodied knowledge). **Explizites Wissen** (explicit knowledge) hingegen ist kommunizierbar, formalisierbar auf verschiedenen Formalisierungsebenen, in verschiedenen Medien speicherbar (Dokumente, Datenbanken, usw.), disembodied knowledge.

Interessant sind die Möglichkeiten der Wissenswandlung und Übertragung:

		Nach	
		implizitem Wissen	explizitem Wissen
Von	implizitem Wissen	Sozialisierung	Externalisierung
	explizitem Wissen	Internalisierung	Kombinierung

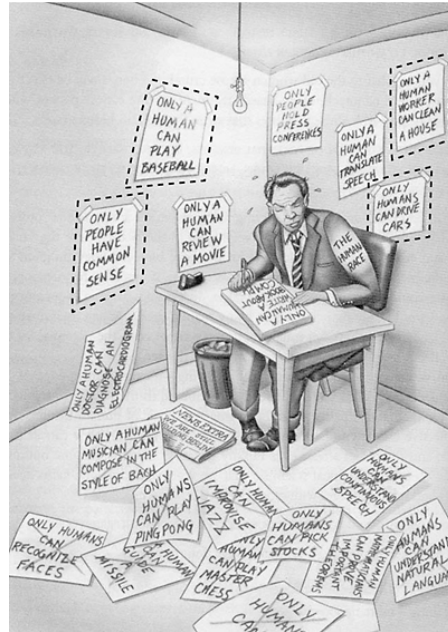
(nach Nonaka & Takeuchi)

Sozialisierung nennt man die Übertragung impliziten Wissens direkt von Person zu Person durch eigene Beobachtung, Nachahmung, praktische Erfahrung, Aufbau eines eigenen mentalen Modells. **Externalisierung** ist die Artikulation von implizitem Wissen durch explizite Konzepte; Verwendung von Metaphern, Analogien, Modellen. **Kombinierung** nennt man die Systematisierung von Konzepten durch Kategorisierung, Sortierung, Kombinierung von Konzepten; Austausch von explizitem Wissen durch Dokumente, Gespräche, ... Schließlich: **Internalisierung** ist die Erzeugung impliziten Wissens aus explizitem Wissen durch learning by doing, zum Aufbau bzw. Erweiterung mentaler Modelle.

Wissen kann in einem Computer oder in einem Computernetz repräsentiert werden. Wissensrepräsentation und Wissensverarbeitung sind die zentralen Themen des Informatik Teilgebietes „**Künstliche Intelligenz**“ (KI, *artificial intelligence* oder *AI*). Alternative Repräsentationen sind z.B.: (künstliche) Neuronale Netze, Semantische Netze, Frames, Begriffliche Graphen, Beschreibungslogiken, Ontologien / Semantic Web.

Ziel der KI-Forschung ist die Entwicklung von Maschinen mit intelligentem Verhalten. Es ist heute üblich, die so genannte „**starke KI**“ (*strong AI*) von der „**schwachen KI**“ (*weak AI*) zu unterscheiden. Die starke KI hat zum Ziel, eine künstliche Intelligenz zu erschaffen, die wie der Mensch **kreativ nachdenken und Probleme lösen kann** und die sich durch eine Form von **Bewusstsein** beziehungsweise Selbstbewusstsein sowie **Emotionen** auszeichnet.

Ein klassischer Prüfstein für das Erreichen dieses Ziels ist der so genannte **Turing Test**. Trotz jahrzehntelangen Bemühens ist dieser immer noch eine der Grand Challenges der Informatik (siehe Vorlesung 0, Die großen Herausforderungen (Grand Challenges) der Informatik). Die KI ist zweifellos eines der Teilgebiete, dessen Ergebnisse am wenigsten überzeugend sind, zumindest gemessen an den Erwartungen und „Versprechungen“:



aus Ray Kurzweil: **Why We Can Be Confident of Turing Test Capability Within a Quarter Century**, Beitrag zum The Dartmouth Artificial Intelligence Conference: The next 50 years (AI@50) on July 14, 2006

Im Gegensatz zur starken KI geht es der **schwachen KI** darum, konkrete (Anwendungs-) probleme zu meistern. Insbesondere sind das solche Probleme, zu deren Lösung nach allgemeinem Verständnis irgendeine Form von „Intelligenz“ notwendig zu sein scheint. Letztlich geht somit um die **Simulation intelligenten Verhaltens** mit Mitteln der Informatik, es geht ihr **nicht** um Schaffung von Bewusstsein oder um ein tieferes Verständnis von Intelligenz.

Kaum ein Teilgebiet der Informatik hat so viele ups-and-downs erlebt, wie die Künstliche Intelligenz: es gab mehrere euphorische Phasen und geradezu Depressionen (z.B. Einstellung der Forschungsförderung). Sie dürfen sich „getrost“ an weiteren Spekulationen z.B. bezüglich des Turing Tests beteiligen. Vielleicht lesen Sie vorher aber z.B. folgende Websites, dass gibt schon etwas mehr Sicherheit:

<http://www.kde.cs.uni-kassel.de/lehre/ws2004-05/wissensverarbeitung>

http://www.iicm.tu-graz.ac.at/Teaching/theses/2000/idb9e_greif/images/node1.html

5 Elemente der Shannonschen Informationstheorie

Vor allem **Claude** Elwood **Shannon** lieferte in den 1940er bis 1950er Jahren wesentliche Beiträge zur Theorie der Datenübertragung und der dazugehörigen Wahrscheinlichkeitstheorie. Er fragte sich, wie man eine verlustfreie Datenübertragung, über elektrische Kanäle, sicherstellen kann. Dabei geht es insbesondere darum, die Datensignale vom Hintergrundrauschen zu trennen. Außerdem versucht man, während der Übertragung aufgetretene Fehler zu erkennen und/oder zu korrigieren.

Die von Shannon formulierte Informationstheorie (1948, A Mathematical Theory of Communication) untersucht Informationsquellen mit statistischen Methoden. Jedes Auftreten eines Zeichens wird dabei als Ergebnis eines Zufallexperimentes. Wenn bei n Beobachtungen (eines Laplace Versuchs) bzw. bei der Überprüfung einer Stichprobe das Ereignis A insgesamt $h_n(A)$ -mal auftritt, dann heißt diese Größe die absolute Häufigkeit des Ereignisses A .

Betrachten wir z.B. die Zeichenfolge "abrakadabra", so fällt auf, dass nicht alle Zeichen (Buchstaben und Freizeichen) gleich häufig vorkommen. Dies gilt für alle natürlichen Sprachen, siehe Anhang 1.

Die folgende Tabelle zeigt die absoluten und relativen Häufigkeiten des Auftretens verschiedener Zeichen für das genannte Beispiel "abrakadabra":

	absolute Häufigkeit $h_n(A) =$ Anzahl des Auftretens im Beispieltext	relative Häufigkeit im Beispieltext $H_n(A) = \frac{h_n(A)}{n}$
a	5	$5/(11 \cdot 100) = 0,45$
b	2	0,18
d	2	0,18
k	1	0,09
r	1	0,09
Summe ($n = \Omega $)	11	1,00

Für große n (also einer großen Stichprobe) wird jedem Ereignis A eine reelle Funktion $P(A)$ als Grenzwert der relativen Häufigkeit $H_n(A)$ zugeordnet, die die Wahrscheinlichkeit des Eintretens des Ereignisses A bestimmt. Also: Wahrscheinlichkeiten kann man durch relative Häufigkeiten schätzen. Relative Häufigkeiten stabilisieren sich, nähern sich dem Grenzwert Wahrscheinlichkeit an, je umfangreicher die Stichprobe ist. Für relative Häufigkeiten und Wahrscheinlichkeiten gilt folgende „Algebra der Ereignisse“:

Ereignis	Interpretation	Relative Häufigkeit	Wahrscheinlichkeit
Ω	Gesamt ereignis (das sichere Ereignis)	$H_n(\Omega) = 1$	$P(\Omega) = 1$
\emptyset	Unmögliches Ereignis	$H_n(\emptyset) = 0$	$P(\emptyset) = 0$

A	Beliebiges Ereignis	$0 \leq H_n(A) \leq 1$	$0 \leq P(A) \leq 1$
$A \cup B$	Es treten die Ereignisse A oder B auf, die „Summe“ von Ereignissen	$H_n(A \cup B) = H_n(A) + H_n(B) - H_n(A \cap B)$	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$A \cap B$	Es treten A und B ein		
$A \setminus B$	Es tritt A und nicht B ein	$H_n(A \cup B) = H_n(A) + H_n(B)$	$P(A \cup B) = P(A) + P(B)$
\bar{A}	Das Ereignis A tritt nicht ein - das komplementäre Ereignis	$H_n(\bar{A}) = 1 - H_n(A)$	$P(\bar{A}) = 1 - P(A)$
$A \subseteq B$	Tritt das Ereignis A ein, dann tritt auch B ein	$H_n(A) \leq H_n(B)$	$P(A) \leq P(B)$
	Die Ereignisse A und B sind unabhängig voneinander	$H_n(A \cap B) = H_n(A) \cdot H_n(B)$	$P(A \cap B) = P(A) \cdot P(B)$

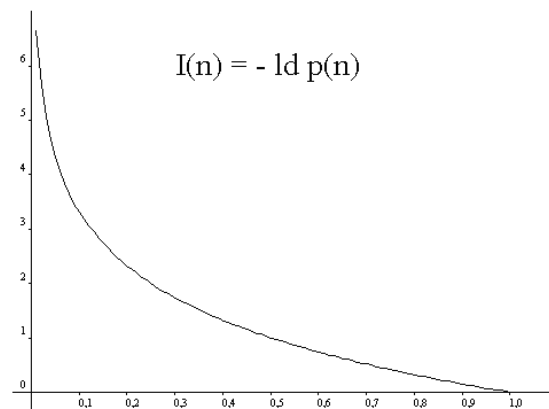
Shannon definiert den **Informationsgehalt einer Nachricht x** als den **negativen Logarithmus dualis** (Logarithmus zur Basis 2) der Wahrscheinlichkeit des Auftretens von x

$$I(x) = -\log_2 P(x)$$

Der Informationsgehalt einer Nachricht hat somit folgende Eigenschaften:

- Mit steigender Wahrscheinlichkeit für das Auftreten einer Nachricht sinkt deren Informationsgehalt.
- Da die Wahrscheinlichkeit einen Wertebereich von 0 bis 1 aufweist, ergibt sich immer ein positiver Wert für den Informationsgehalt einer Nachricht.
- Für Nachrichten mit gegen Null gehender Wahrscheinlichkeit steigt der Informationsgehalt stark an und geht für $P(x) \rightarrow 0$ gegen unendlich.
- **Bei voneinander unabhängigen Nachrichten addiert sich der Informationsgehalt.**
- Der Informationsgehalt gibt die günstigste (kürzeste) Codelänge für die Darstellung des Zeichens in Bit an.
- Wenn die Wahrscheinlichkeit für das Auftreten eines Zeichens den Wert 0,5 annimmt, ist der Informationsgehalt gleich 1. Ein idealer Code würde für dieses Zeichen eine Codelänge von einem Bit verwenden.

Graphisch stellt sich der Zusammenhang zwischen Auftrittswahrscheinlichkeit und Informationsgehalt folgendermaßen dar.



Betrachtet man jetzt eine Nachrichtenquelle (Datenquelle) dann gilt: Eine Quelle hat eine hohe **Entropie**, wenn deren Inhalte (Nachrichten) eine eher zufällige Verteilung aufweisen. Die Entropie sinkt, wenn bestimmte Nachrichten häufiger auftreten als andere. Mit Hilfe des Entropiebegriffs lässt sich bestimmen, wie gut (hoch) die Ausgabe einer Quelle komprimiert werden kann, ohne dass Informationen verloren gehen. Bei bekannter Wahrscheinlichkeitsverteilung gibt die Entropie einer Nachrichtenquelle an, mit welcher durchschnittlichen Kodelänge (in Bit) diese kodiert werden kann.

Shannon definierte die **Entropie H** einer Nachrichtenquelle (Informationsquelle) I über einem Zeichenvorrat Ω durch

$$H(I) = - \sum_{j=1}^{|\Omega|} p_j \cdot \log_2 p_j$$

wobei p_j die Wahrscheinlichkeit ist, mit der das j -te Zeichen des Zeichenvorrats Ω im Informationstext I auftritt. Die **Entropie** erhält die **Einheit bit**.

Die Entropie einer Nachrichtenquelle ist also ausschließlich von der Wahrscheinlichkeit p_j abhängig, mit der die einzelnen Nachrichten auftreten. Um die Entropie zu berechnen, muss für alle möglichen Nachrichten einer Quelle die Wahrscheinlichkeit bekannt sein, mit der diese auftreten.

Beispiel: Nehmen wir an, dass die Häufigkeit des Auftretens einzelner Zeichen in dem Wort "abrakadabra" repräsentativ für die Quelle ist, so ist

Zeichen A	relative Häufigkeit im Beispieltext $H_n(A) = \frac{h_n(A)}{n}$	$-p(A) \log_2 p(A)$ $= -H_n(A) \log_2 H_n(A)$
a	$5/(11 \cdot 100) = 0,455$	0,517
b	0,182	0,447
d	0,182	0,447
k	0,091	0,314
r	0,091	0,314
Summe ($n= \Omega $)	1,00	2,040

In diesem Beispiel beträgt die Entropie der Nachrichtenquelle 2,04 bit.. Theoretisch ist bei dieser Häufigkeitsverteilung eine Kodierung der 5 verschiedenen Zeichen mit 2,04 Bit pro Zeichen erreichbar, das wäre eine Optimalcodierung. Das heißt mit anderen Worten: Es gibt keine Codierung dieser Informationsquelle, die mit zwei Bit auskäme, man braucht leider etwas mehr. Man kann sich dem Wert 2,04 aber praktisch annähern, indem man verschieden lange Codewörter benutzt: für die häufig auftretenden Zeichen kurze Codewörter, für die seltener auftretenden Zeichen entsprechend längere Codewörter, etwa nach der Huffman-Kodierung. Bevor wir ein Konstruktionsschema kennen lernen, wollen wir uns erst mit den Eigenschaften der Entropie vertraut

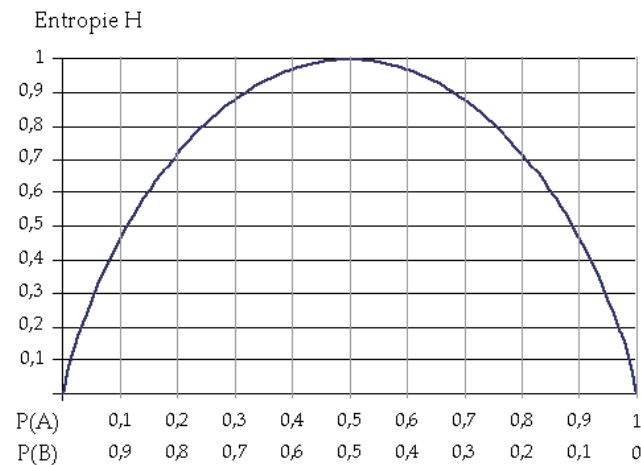
Schauen wir uns die Formel $H(I) = - \sum_{j=1}^{|\Omega|} p_j \cdot \log_2 p_j$ etwas genauer an:

Minimalwerte: Sofern immer nur eine bestimmte Nachricht eintrifft, also die Wahrscheinlichkeit für deren Auftreten gleich 1 ist, erreicht die Entropie einer Nachrichtenquellen ihren Minimalwert. Es ergibt sich ein Betrag von $(1 \text{ ld } 1) \rightarrow 0$. Die Nachrichtenquelle stellt keinerlei Information zur Verfügung, es besteht keine Unsicherheit über die vorraussichtlich nächste Nachricht.

Wenn eine bestimmte Nachricht von der Nachrichtenquelle überhaupt nicht ausgesendet wird, ergibt sich für diese Nachricht ein Betrag von $(0 \text{ ld } 0) \rightarrow 0$, d.h. auch sie trägt nicht zur Information auf Seiten des Empfängers bei.

Maximalwerte: Die Entropie einer Nachrichtenquelle erreicht ihren Maximalwert, wenn alle Nachrichten mit der gleichen Wahrscheinlichkeit auftreten. In diesem Fall besteht die größtmögliche Unsicherheit über die Natur der zu erwartenden Nachricht. Folglich ist bei deren Eintreffen der Informationsgewinn am größten.

Funktionsverlauf: Entropie einer **binären** Nachrichtenquelle: In dem nachstehenden Diagramm ist der Verlauf der Entropie einer Nachrichtenquelle dargestellt, die zwei unterschiedliche Nachrichten (A und B) aussenden kann:



Für die Wahrscheinlichkeit dieser beiden Nachrichten gilt immer

$$P(A) = 1 - P(B),$$

da die Summe beider Wahrscheinlichkeiten zwangsläufig den Wert 1 ergeben muss.

Der Maximalwert für die Entropie wird mit einem Betrag von 1 erreicht, wenn beide Nachrichten mit identischer Wahrscheinlichkeit auftreten (je 0,5). In diesem Fall müsste die Nachrichtenquelle mit einer durchschnittlichen Codelänge von 1 Bit kodiert werden. Weicht die Wahrscheinlichkeitsverteilung in die eine oder andere Richtung davon ab, so besteht theoretisch die Möglichkeit einer günstigeren Kodierung.

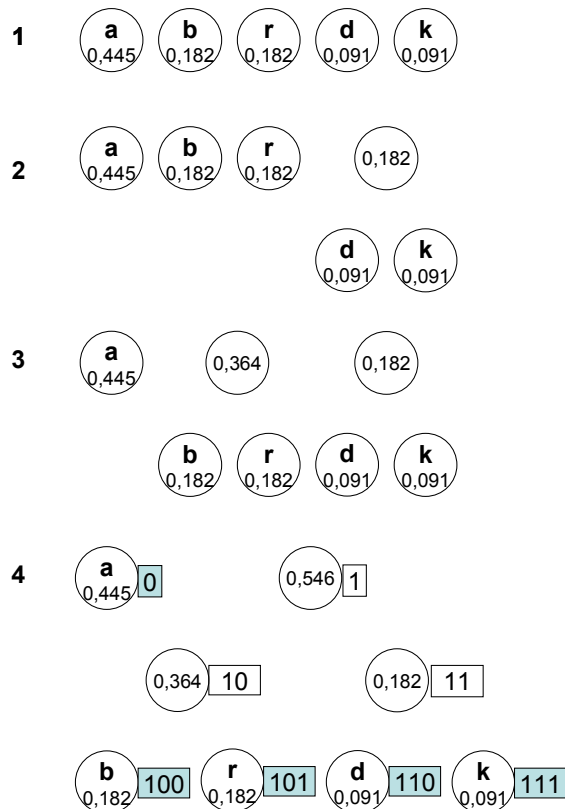
Zwischen Entropie und (bedeutsamer) Information (Informationsgehalt)- besteht ein nichtmonotoner Zusammenhang. Ist die Entropie gleich Null (regelmäßiges, 100%-ig vorhersagbares Signal), so ist der (potenzielle) Nutzwert für einen Organismus genauso gering wie bei der gegen unendlich strebenden Entropie eines Rauschsignals. Das Informations-Maximum liegt also, grob gesprochen, in einem "mittleren" Entropiebereich.

Wenn die Entropie etwa einen Wert von 1 hat, dann gilt die Information als zufällig. Bei einer kleinen Entropie enthält der Informationstext Redundanzen oder **statistische** Regelmäßigkeiten. Die Zahl $H(I)$ gibt intuitiv die durchschnittliche Information an, die in einem Zeichen der Quelle enthalten ist.

Wir wollen jetzt kurz unsere Erkenntnisse erstmals praktisch anwenden. Hierzu betrachten wir noch einmal unser Beispiel "abrakadabra". Wir haben uns versichert, dass es keinen Code mit einer durchschnittlichen (d.h. mit der Wahrscheinlichkeit gewichteten) Codelänge von weniger als 2,04 Bit geben kann. Kanonisch würden wir für die 5 Zeichen also 3 Bit benötigen, wie in der Tabelle unten angegeben. Diesen Unterschied zwischen (mittleren) Codewortlänge und Entropie nennen wir **Redundanz**, in unserem Fall $R = 3 \text{ bit} - 2,04 \text{ bit} = 0,96 \text{ bit}$.

Zeichen A	relative Häufigkeit im Beispieltext $H_n(A) = \frac{h_n(A)}{n}$	$-p(A) \log_2 p(A)$ $= -H_n(A) \log_2 H_n(A)$	kanonischer Binär-code	Huffman Code	Gewichtete Codelänge $p(A) \cdot C_n $
a	$5/(11 \cdot 100) = 0,455$	0,517	000	0	0,455
b	0,182	0,447	001	100	0,545
d	0,182	0,447	010	110	0,545
k	0,091	0,314	011	111	0,273
r	0,091	0,314	100	101	0,273
Summe ($n= \Omega $)	1,00	2,040 bit	3 bit		2,091 bit

Wir können allerdings einen Code mit **variabler Codelänge** konstruieren (der natürlich trotzdem eindeutig dekodierbar sein muss), der im Mittel deutlich weniger Übertragungsaufwand bedeutet. Ein Beispiel hierfür ist der **Huffman Code**. Dieser lässt sich nach folgendem Schema konstruieren:



Schritt 1: Wir notieren die zu kodierenden Zeichen mit ihren Häufigkeiten als Knoten.

Schritt 2: Wir fassen die zwei Knoten (allgemeiner zwei Knoten) mit den geringsten Häufigkeiten zusammen zu einem „virtuellen“ Knoten (Bedeutung hier d oder k), der mit der Summenhäufigkeit auftritt.

Schritt 3: Wie Schritt 2

Schritt 4: Wie Schritt 2; Abbruch, wenn auf oberster Ebene nur noch zwei Knoten vorhanden sind.

Abschluß: Wir ordnen allen Knoten Codewerte zu:

- Auf 1. Ebene 0 und 1
 - 2. Ebene 10 und 11
 - 3. Ebene 100, 101, 110 und 111
- jeweils, soweit noch „Nachfolger“ vorhanden sind.

Auf diese Art wird jedem Zeichen ein Codewort zugeordnet, es entsteht die oben angegebene Code-Tabelle. Die Analyse der mit der Häufigkeit des Auftretens gewichtete Codelänge zeigt, dass wir sehr nahe am theoretischen Optimum landen: Wir haben eine mittlere Codelänge von 2,091 Bit gegenüber dem Optimum 2,040, also lediglich eine Redundanz von 0,051 bit und das ist zu vergleichen mit dem brute-force Ergebnis von 3 Bit und einer Redundanz von 0,96 bit. Der Huffman Code ist ein Beispiel für die so genannte **Entropiecodierung**.

Die rein statistische Berechnung der informationstheoretischen Entropie nach obiger Formel ist gleichzeitig ihre Beschränkung. Beispielsweise ist die Wahrscheinlichkeit, eine 0 oder 1 in einer regelmäßigen Zeichenkette "1010101010..." zu finden, genauso groß, wie in einer Zeichenkette, die durch statistisch unabhängige Ereignisse (etwa wiederholten Münzwurf) entstanden ist. Daher ist Shannons Entropie für beide Zeichenketten identisch, obwohl man intuitiv die erste Kette als weniger zufällig bezeichnen würde. Eine angemessenere Definition der Entropie einer Zeichenkette liefert die **bedingte Entropie** und **Quellentropie**, die beide auf Verbundwahrscheinlichkeiten aufbauen.

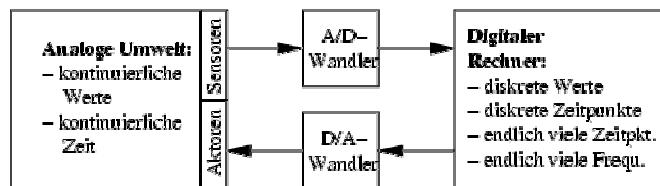
8

⁸ Klassische Informationskonzepte versagen teilweise in quantenmechanischen Systemen. Dies führt zum Konzept der Quanteninformation.

6 Diskretisierung (Abtastung) und Quantisierung

6.1 Grundsätzliches

Bei vielen Computeranwendungen steht die digitale Verarbeitung analoger Signale, oder kurz: Digitale Signalverarbeitung im Vordergrund. Mit “analog” meinen wir zeit- und wertkontinuierliche Größen, wie sie in vielen mechanischen, optischen und elektrischen Systemen anzutreffen sind. Mit “digital” bezeichnen wir zeit- und wertdiskrete Systeme, wie zum Beispiel einen Digitalrechner.



Zur digitalen Verarbeitung analoger Signale benötigt man - wie in obiger Abbildung skizziert - Schnittstellen zwischen der “analogen”, also zeit- und wertkontinuierlichen, Umwelt und einem “digital”, also zeit- und wertdiskret arbeitenden, System, in der Regel einem Mikroprozessor. Diese Schnittstellen zwischen analoger Umwelt und digitalem Prozessor arbeiten in zwei Richtungen:

Sensoren wandeln physikalische Größen in eine analoge, elektrische Spannung. Die Spannung ist in der Regel zeit- und wertkontinuierlich. Die Aufgabe eines **Analog/Digital-Wandlers** (engl. *Analog/Digital converter*, ADC) ist es, diese Spannung in eine zeit- und wertdiskrete, digitale Darstellung zu wandeln.

Ausgaben des Mikroprozessors werden als digitale Werte an einen **Digital/Analog-Wandler** (engl.: *Digital/Analog Converter*, DAC) ausgegeben. Dieser berechnet aus zeit- und wertdiskreten Signalen, welche der Prozessor liefert, zeit- und wertkontinuierliche Signale, die über Aktoren physikalische Größen verändern. Aus digitalen Signalen müssen analoge Signale rekonstruiert werden können.

Die **Digitalisierung** eines analogen Signals erfolgt in zwei Schritten.: **Abtastung** und **Quantisierung**. Dadurch entsteht ein zeit- und wertdiskretes Signal, das in einem Digitalcomputer verarbeitbar ist. Anders herum bezeichnet man die Wandlung eines diskreten, quantisierten Signals in ein kontinuierliches Signal als **Rekonstruktion**.

Abtastung (englisch *sampling*) ist die Registrierung von Messwerten zu diskreten, häufig äquidistanten Zeitpunkten. Aus einem **zeitkontinuierlichen** Signal wird so ein **zeitdiskretes** Signal gewonnen, ein erster Schritt um Werte der realen Welt im Computer zu behandeln.

Die Anzahl der Abtastungen pro Sekunde wird Abtastrate genannt. Bei der digitalen Telefonie (ISDN) z.B. beträgt die Abtastrate 8 kHz. Digitales Video (professionelle Formate) wird dagegen mit 13,5 MHz abgetastet.

6.2 Modell der Abtastung

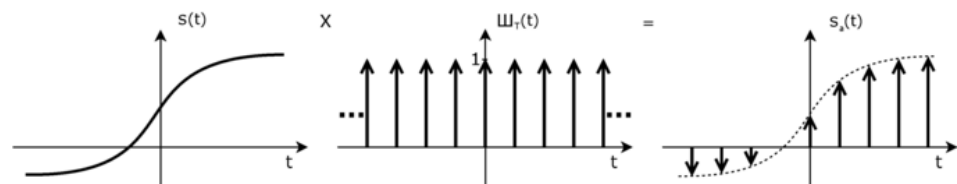
Das Grundprinzip der Abtastung ist, dass eine physikalische Größe (z.B. ein Signal in Form einer elektrischen Spannung) über einen gewissen Zeitraum akkumuliert wird und dann der

akkumulierte Wert in eine digitale Darstellung des Meßwertes umgewandelt wird. Dies geschieht dann in der Regel periodisch mit der Abtastfrequenz

Ideale Abtastung

Für eine einfachere mathematische Beschreibung ist die ideale Abtastung definiert. Hier wird das Signal nicht über einen gewissen Zeitraum um den Abtastzeitpunkt akkumuliert, sondern exakt zum Abtastzeitpunkt nT ausgewertet.

Mathematisch lässt sich dies darstellen, indem man das Signal $s(t)$ mit dem Dirac-Kamm (einer Folge von Dirac-Stößen) multipliziert:



Das abgetastete Signal lautet dann

$$s_a(t) = s(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

Das **(Nyquist-)Shannonsche Abtasttheorem**, manchmal auch (WKS-Sampling-Theorem, für Whittaker-Kotelnikow-Shannon) genannt, ist ein grundlegendes Theorem der Nachrichtentechnik, Signalverarbeitung und Informationstheorie. Es war auch Claude Elwood Shannon der es 1948 als Ausgangspunkt seiner Theorie der maximalen Kanalkapazität formulierte, d.h. der maximalen Bitrate in einem frequenzbeschränkten, rauschbelasteten Übertragungskanal (für Shannon damals nur ein Lemma!) ⁹

Das Abtasttheorem besagt, dass ein kontinuierliches, **bandbegrenzte**s Signal mit einer Minimalfrequenz von 0 Hz und einer Maximalfrequenz f_{\max} mit einer Frequenz größer als $2 \cdot f_{\max}$ abgetastet werden muss, damit man aus dem so erhaltenen zeitdiskreten Signal das Ursprungssignal ohne Informationsverlust (aber mit unendlich großem Aufwand) rekonstruieren bzw. (mit endlichem Aufwand), das heißt beliebig genau approximieren kann, also

⁹ Dabei stützte er sich auf Überlegungen von Harry Nyquist (1928) zur Übertragung endlicher Zahlenfolgen mittels trigonometrischer Polynome und auf die *Theorie der Kardinalfunktionen* von Edmund Taylor Whittaker (1915) und seinem Sohn John Macnaughten Whittaker (1929) [2]. Unabhängig davon wurde das Abtasttheorem 1933 von Wladimir Alexandrowitsch Kotelnikow [3] in der sowjetischen Literatur eingeführt, was im Westen allerdings erst in den 1950er Jahren bekannt wurde. Ansätze zur Interpolation mittels Kardinalreihen oder ähnlicher Formeln lassen sich bis in die Mitte des 19. Jahrhunderts zurückverfolgen.

$$f_{\text{abtast}} > 2 f_{\text{max}}$$

In der Praxis bedeutet das Abtasttheorem, dass man vor der Abtastung die maximale Frequenz kennen oder herausfinden muss (zum Beispiel mit Hilfe der Fourier-Analyse und dass dann das Signal (zum Zwecke der Digitalisierung) mit mehr als der doppelten Frequenz abgetastet werden muss, wenn man das Signal in guter Näherung rekonstruieren will. Mißachtet man dieses, so tritt **Aliasing** auf, das sind ggf. schwerwiegende Fehler, die erheblich Effekte zeigen!

f_{max} nennt man, nach Vorschlag von C. E. Shannon, die **Nyquist-Frequenz**.

Analog gilt das Abtasttheorem auch bei Bildern, wobei die Abtastfrequenz dann in Linien (bzw. Pixel) pro Längeneinheit bestimmt werden kann, aus Zeitfrequenz wird Ortsfrequenz!

Bemerkung: Nur wenige reale Signale erfüllen die Voraussetzungen eine bandbeschränkte Funktion zu sein, z.B. Signale mit Knicken oder Sprüngen. gehören nicht dazu. Das Abtasttheorem ist somit als ideale Aussage in einer idealen Situation zu betrachten. Dem Ideal am nächsten kommen modulierte Schwingungen, wie Musik- oder Sprachaufzeichnungen, welche zum Speichern auf CD gesampelt und digitalisiert werden sollen. Für andere praktische Zwecke, z.B. digitale Bildbearbeitung, müssen Varianten des Abtasttheorems mit nicht ganz so starken Anforderungen gefunden werden, für welche dieses Theorem dann Richtschnur ist.

Dieses ist aber nur der erste Teil des Abtasttheorems, der zweite ist konstruktiv und sagt, wie ein Signal rekonstruiert werden kann, aber das würde hier zu weit führen.

Für die Verarbeitung in Digitalrechnern ist ein zweiter Vorgang wichtig, nämlich die Quantisierung. Der Begriff **Quantisierung**, nicht zu verwechseln mit Quantifizierung, kommt vom *lat. quantum* (zu Deutsch: "Menge") und bedeutet etwa "Darstellung einer Größe in einem System, in dem sie nur **diskrete** Werte annehmen kann, d.h. in der Regel auch nur endlich viele.

7 Datenverarbeitung, Informationsverarbeitung, Wissensverarbeitung

Kehren wir wieder zur ursprünglichen Betrachtung, dem Zusammenhang zwischen Daten und Informationen zurück.

Eine Menge **D** von Daten D_i ist von Interesse, wenn ihr (mindestens) eine Menge **I** von Informationen I_i und mindestens eine (Menge von) Abbildungen

$$D_i \xrightarrow{\alpha} I_i$$

entspricht., welche wir Interpretationen nennen. Wir schreiben hierfür

$$\mathbf{D} \xrightarrow{\alpha} \mathbf{I}$$

Da die Datenmenge \mathbf{D}' ebenfalls einer Informationsmenge \mathbf{I} entspricht erhalten wir mit einer Datenverarbeitungsvorschrift

$$\mathbf{D} \xrightarrow{\nu} \mathbf{D}'$$

den folgenden Zusammenhang:

$$\begin{array}{ccc} \mathbf{D} & \xrightarrow{\alpha} & \mathbf{I} \\ \downarrow \nu & & \downarrow \sigma \\ \mathbf{D}' & \xrightarrow{\alpha'} & \mathbf{I}' \end{array}$$

Welche Beziehung besteht jetzt zwischen \mathbf{I} und \mathbf{I}' ?

Sicher entspricht jedem Datum D ein Paar (I, I') , mit $I = \alpha(D)$, $I' = \alpha(D')$ also eine Zuordnung σ zwischen I und I' . Ist α nicht umkehrbar, gibt es also zwei Daten D_1 und D_2 , welche die gleiche Information I wiedergeben, so braucht die Zuordnung σ keine Abbildung zu sein!

Wir nennen eine Verarbeitungsvorschrift ν **informationstreu**, wenn σ eine Abbildung ist. σ heißt dann eine Vorschrift zur **Informationsverarbeitung**.

Gewöhnlich werden Daten überhaupt nur bearbeitet, um eine bestimmte Informationsverarbeitung zu erzielen. Man geht also in der Regel davon aus, dass man eine bestimmte Informationsverarbeitung σ erzielen will und versucht α, α', ν so zu bestimmen, dass eine Informationsverarbeitung realisiert wird. Was sind die Bedingungen?

Diese Zusammenhänge werden noch vielfach wiederkehren, zum Beispiel im Zusammenhang mit Datentypen, abstrakten Datentypen, Objekten.

Wissensverarbeitung (*knowledge engineering*) beschäftigt sich mit der organisatorischen und technischen Unterstützung von Wissensprozessen. Wichtige Aktivitäten der Wissensverarbeitung sind das Entdecken und Strukturieren von Wissen, das Ableiten von neuem Wissen, und die Kommunikation des Wissens. Zentraler Punkt all dieser Aktivitäten ist die Frage, wie das Wissen im Computer repräsentiert werden kann.

Sofern es sich um anspruchsvolles Wissen handelt, spielt bei all diesen Aktivitäten der Mensch eine zentrale Rolle. Da das menschliche Denken begrifflich organisiert ist, entstand eine Reihe von Forschungsgebieten, die eine Semantik-basierte Unterstützung der Aktivitäten zum Thema haben: Knowledge Discovery, Ontologien/Metadaten, Semantic Web, Formale Begriffsanalyse, sowie Visualisierung und Interaktion.

8 Gesellschaftliche Wirkungen

Seit den 70er Jahren des 20. Jahrhunderts kristallisieren sich zur Kennzeichnung unserer Gesellschaft zwei „neue“ Begriffe heraus: Informationsgesellschaft und/oder Wissensgesellschaft und stelle diese in die historische Entwicklungslinie von der Agrargesellschaft über die Industriegesellschaft. Beide Begriffe kennzeichnen eine Gesellschaft in der

- Wissen und/oder Information eine strategische Ressource in Produkten und Dienstleistungen ist
- das Wissen/die Information zunehmend vernetzt, dezentral und interdisziplinär ist
- die effektive Nutzung des Wissens/der Information entscheidender Wettbewerbsfaktor geworden ist und
- Wissen und Information selbst zum veräußerlichen Gut wird.

Informationen, Wissen und Bildung werden als wichtigste **Rohstoffe** unserer Zeit bezeichnet.

Der oben gemachten Unterschied zwischen Information und Wissen wird in diesen Begriffsbildungen allerdings nicht aufgegriffen, vielmehr bezeichnet der Begriff **Informationsgesellschaft** das Leitbild einer auf Informations- und Kommunikationstechnologien basierenden postindustriellen oder postmodernen Gesellschaft und Ökonomie. Verschiedene Autoren betonen verschiedene Aspekte

- Ökonomischen Aspekte und einhergehende wirtschaftlichen Veränderungen, z.B. Herausdifferenzierung eines "quartären" Sektors¹⁰ (Dienstleistungen mit Informationen);
- IuK-Technologien als wesentlichen Faktor der wirtschaftlichen (und gesellschaftlichen) Entwicklung;
- Nutzungsaspekts und der Bedeutung für die Menschen in einer Informationsgesellschaft; auch "*informierte Gesellschaft*" (Steinbuch 1966), "*informationsbewusste Gesellschaft*" (Wersig 1973).

Bereits 1948 prognostizierte der amerikanische Wissenschaftler Norbert Wiener¹¹ die Entstehung einer "Informationsgesellschaft".

¹⁰ In Erweiterung der Drei-Sektoren-Hypothese für die wirtschaftliche Entwicklung eines Staates nach **Jean Fourastié** (französischer Ökonom): Der **Primärsektor**, auch Urproduktion liefert zumeist die Rohstoffe für ein Produkt (Landwirtschaft, Forstwirtschaft, Fischerei und Bergbau). Der **Sekundärsektor**, auch industrieller Sektor umfasst das produzierende Gewerbe, das heißt die Verarbeitung von Rohstoffen, also z.B. Verarbeitendes Gewerbe, Industrie, Handwerk, Energiewirtschaft: Der **Tertiärsektor** umfasst den gesamten Bereich der Dienstleistungen.

¹¹ **Norbert Wiener** (1894-1964) war ein amerikanischer Mathematiker. Er wurde als Begründer der Kybernetik bekannt, ein Ausdruck, den er in seinem Werk *Cybernetics or Control and Communication in the Animal and the Machine* (1948) prägte. Heute versteht man unter Kybernetik die Wissenschaft von der Struktur komplexer Systeme, insbesondere der Kommunikation und Steuerung einer Rückkopplung (engl. Feedback) bzw. eines Regelkreises. Der Begriff Kybernetik war aber schon viel früher in Gebrauch: Homer

Viele Autoren bemühen sich, den auf technische und wirtschaftliche Aspekte reduzierten Begriff „Informationsgesellschaft“ weiter zu fassen und bevorzugen den angeblich flexibleren und weiteren Begriff der **Wissensgesellschaft**:

„Die Wissensgesellschaft eröffnet eine Perspektive, die auf den Willen und die Befähigung der Menschen zu Selbstbestimmung setzt - ganz im Gegensatz zum technizistischen Begriff der Informationsgesellschaft.“

Nicht Rechnerleistungen und Miniaturisierung werden die Qualität der künftigen gesellschaftlichen Entwicklung bestimmen. Entscheidend wird die Auswahl des Nützlichen und die Fähigkeit zum Aushalten von Ambivalenzen und Unsicherheit sein, die Gestaltung des Zugangs zu Wissen und der fehlerfreundliche Umgang mit dem Nichtwissen. Wissen wird zur Schlüsselressource, Bildung zur Bedingung für die Teilhabe am gesellschaftlichen Leben.“ (<http://www.wissensgesellschaft.org/>, Heinrich Böll Stiftung)

Die Bezeichnung „Wissensgesellschaft“ wird vor allem im geistes- und sozialwissenschaftlichen, (bildungs-)politischen und privatwirtschaftlichen Zusammenhang gebraucht, um verschiedene gesellschaftliche Entwicklungen, die aus einer Neugewichtung von intellektuellem und praktischem Wissen entstanden sind, zu erfassen. Ob damit eine im Entstehen begriffene oder tatsächlich bereits real existierende Gesellschaftsform beschrieben ist, bleibt jedoch streitig.

schrieb "kybernètès" und meinte den Steuermann eines Schiffes. Platon benutzte diesen Begriff wenn er vom "Mann am Steuerruder einer Regierung" sprach. Der Apostel Paulus artikuliert sich mit dem Begriff "kybérnēsis" in 1.Korinther 12:28 um die "Fähigkeiten zu leiten" zu thematisieren.