



# 1 Einführung

## 1.1 Einleitung

---

Eine Definition von Daten durch die ZHAW Studierenden 2016

**Samuel Reutimann:** *Data that is accurate and timely, specific and organized for a purpose, presented within a context that gives it meaning and relevance, and can lead to an increase in understanding and decrease in uncertainty.*

*Information is valuable because it can affect behavior, a decision, or an outcome. A piece of information is considered valueless if, after receiving it, things remain unchanged.*

**Bernhard Fuchs:** *Information ist eine Teilmenge an Wissen, die ein Sender einem Empfänger mittels Signalen über ein bestimmtes Medium (auch ‚Informationskanal‘ genannt) vermitteln kann.*

*Andere Definitionen sind oft stark vom Verwendungskontext des Begriffs «Information» abhängig oder zu sehr auf das IT-Umfeld bezogen.*

**Marco Buess:** *Informationen bilden im Besonderen den Inhalt und den Bedeutungsgehalt einer Nachricht, in textlicher, grafischer oder audiovisueller Form. Informationen enthalten keine irrelevanten oder redundanten Teile. Information ist das Wissen über Funktionen, Arbeitsweisen, Verfahren und Vorgänge. Diese Wissens-Ressource wird in modernen Unternehmen für die Realisierung der strategischen Unternehmensziele benutzt und durch das Informationsmanagement in vorhandene und neue Anwendungen eingebracht.*

*Informationen können durch Daten dargestellt und auf Datenträgern gespeichert, in Computern verarbeitet und über Ausgabegeräte ausgegeben werden. Andererseits lassen sich Informationen aus Daten reproduzieren, wenn sie entsprechend interpretiert werden können.*

**Daniel Brun:** *Information: Human-consciousness-level symbols needing a context to give them meaning. For example, the price of an airline ticket or a book.*

*Information: the degree of order or complexity of a system.*

**Manuel Martin:** *Information as a representation of knowledge* Information is stored knowledge. Traditionally the storage medium has been books, but increasingly electronic media are becoming important.

*Information as data in the environment* Information can be obtained from a range of environmental stimuli and phenomena not all of which are intended to 'convey' a message, but which can be informative when appropriately interpreted.

*Information as part of the communication process;* Meanings are in people rather than in words or data. Timing and social factors play a significant role in the processing and interpretation of information.

*Information as a resource or commodity* Information is transmitted in a message from sender to receiver. The receiver interprets the message as intended by the sender. There may be added value as the information is disseminated or exchanged.

**Simon Lang:** *Derjenige Anteil einer Nachricht, der für den Empfänger einen Wert besitzt. Durch Informationen werden beim Wirtschaftssubjekt bestehende Wahrscheinlichkeitsurteile bez. entscheidungsrelevanter Daten oder Ereignisse (z.B. Tauschmöglichkeiten oder technische Innovationen) verändert.*

Information (lateinisch: informare = bilden, durch Unterweisung Gestalt geben) ist der abstrakte Grundstein der Moderne. Es wird von einer Informationsgesellschaft gesprochen, wenn unsere heutige Gesellschaft gemeint ist. Sie unterscheidet sich durch die Verbreitung der Informationstechnologie von anderen Zeitabschnitten, wie zum Beispiel dem Maschinenzeitalter. Trotz der alltäglichen Nutzung der Informationstechnologie, sind fundamentale Fragen bezüglich der Ware Information nicht geklärt. So ist die Frage, wie viel eine Information kostet, nicht geklärt. Der Wert einer bestimmten Information kann nicht, wie andere Unternehmenswerte, aufgrund von Bewertungsgrundlagen oder Buchhaltungsregeln berechnet werden. Auch wird Information auf sehr viele Arten definiert und wir wissen leider noch nicht genau, wie Menschen mit Informationen umgehen, also welche Prozesse im Hirn genau stattfinden. Was jedoch bekannt ist, wie Maschinen mit Informationen umgehen, wie Maschinen Informationen interpretieren. Jede Komponente, die in der Informationstechnologie eingesetzt wird, ob nun Hardware oder Software, stellt eine solche Maschine dar. Diese Vorlesung stellt einen Versuch dar, die maschinelle Verarbeitung von Informationen aus verschiedenen Blickwinkeln darzustellen. Die Themenauswahl umfasst die kognitive Psychologie, Browser, Parser und Compiler, Prozessoren und Converter, Data Lifecycle Management, Petabyte Computing und andere Aspekte der Interpretation von Informationen durch die Informatik. Die Auswahl ist subjektiv und umfasst nicht den Grundlagenstoff einer Informatikausbildung, sondern Aspekte, die vielleicht nicht auf der Hand liegen, die jedoch in der täglichen Arbeit sehr nützlich zu wissen sind. Für eine Reihe von Lösungen ist dieses Wissen sogar unentbehrlich. Es soll Informatikfachleuten helfen, über den Tellerrand der täglichen Arbeit zu blicken und ab und an eine etwas bessere Lösung zu finden.

### 1.1.1 Ein Blick in die nahe Zukunft

Niemand sagte es schöner als der CEO von Xing, Stefan Gross-Selbeck: „*Daten sind das Öl des 21. Jahrhunderts.*“ So lautet das Schlagwort der Stunde denn auch: „Big Data“. Unter diesem Begriff werden sowohl die Daten und deren spezielle Eigenschaften als auch die innovativen Technologien zur Analyse dieser Daten verstanden. Anders gesagt: Hinter dem Begriff „Big Data“ steckt nichts anderes, als das systematische Fördern dieses Rohstoffes des 21. Jahrhunderts mit modernster Technologie.

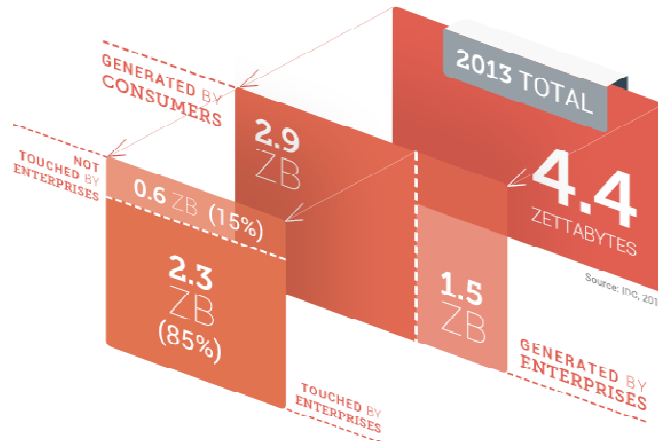


Abbildung 1.1 Datenaufkommen 2013 gemäss EMS und IDC

Die Menge dieser Daten wächst sehr stark. Die interaktive Studie „Digitales Universum“, die im letzten Jahr gemeinsam von EMC und IDC veröffentlicht wurde, sagt für das Jahr 2015 ein Datenaufkommen von über 8 Zettabytes voraus – 40% mehr als im Jahr davor. Bis zum Jahr 2020 sollen es 44 Zettabytes sein. Und der überwältigende Anteil dieser Daten kommt aus dem Netz der Sozialen Medien, wird von Sensoren und anderen Geräten erzeugt und als Bild, Film, Ton, Office-Dokumente oder in anderen Formaten gespeichert. Speziell an diesen Daten ist, dass sie in immer grösserer Geschwindigkeit erzeugt werden, in immer grösserer Vielfalt vorliegen und dass sie nicht immer zu 100% verifizierbar sind, beispielsweise im Falle von fehlenden, falschen oder doppeldeutigen Daten. Neue Analyseverfahren sollen es erlauben, aus dem Datenmeer Informationen zu gewinnen, die Unternehmen helfen, den Kunden und den Markt besser zu verstehen, Produkte und Services schneller und gezielter zu platzieren, Kosten zu sparen, Umsätze zu steigern und Wettbewerbsvorteile auszubauen – also schlicht und einfach, die unternehmerischen Ziele rascher zu erreichen.

„*Daten bleiben scheu und grausam. Es ist schwierig, aus ihnen zu lernen, und noch schwieriger, das Gelernte umzusetzen*“, sagte der Data-Mining-Pionier Nicolas Bissantz bereits vor zwei Jahren im Wissensmagazin des Gottlieb Duttweiler Instituts. Konkret bedeutet dies, dass nicht aus jeder technisch ableitbaren Wechselbeziehung auch eine wirtschaftlich relevante Ursache abgeleitet werden kann. Die Auswertung von Daten ist und

bleibt eine aufwändige Arbeit und erfordert grosses Fachwissen. Aus diesem Grund hat sich ein neuer multidisziplinärer Ansatz für die Aufbereitung, die Bereitstellung und den Zugang zu Daten, Analysen und Auswertungen entwickelt, der sich „Data Science“ nennt. Viele Unternehmen ergänzen bereits heute ihre Analyseteams mit Data-Science-Spezialisten, um aus dem Rohstoff Daten möglichst schnell hochwertige Informationen zu gewinnen. Grössere Unternehmen werden sogar auf ganze Data-Science-Teams setzen.

## 1.2 Was ist Information

---

Die sehr oft in der Informatik verwendete Definition von Information ist diejenige von Shannon in seinem Artikel "A Mathematical Theory of Communication" aus dem Jahre 1948 [Shannon 1948]. Darin definiert er Information als statistisches Mass der Ungewissheit bei der Übertragung einer Nachricht, also im weitesten Sinne als Signal. Je höher die Wahrscheinlichkeit, dass eine bestimmte Information bereits vorhanden ist, desto kleiner der Informationswert der Nachricht. Shannon trennt streng zwischen Bedeutung und Information, die seiner Ansicht nach nichts miteinander zu tun haben. Gemäss verschiedenen Quellen soll er diese Auffassung bereits 1940 entwickelt haben. Er hat jedoch, wie viele amerikanische Wissenschaftler dieser Zeit, die eine Basis für die moderne Informationstechnologie geschaffen haben, mit der Veröffentlichung bis nach dem zweiten Weltkrieg warten müssen.

Neben der Informatik befassen sich die Naturwissenschaften und die Sprachwissenschaften aber auch die Philosophie und andere Bereiche der Wissenschaft mit dem Phänomen Information. Es existiert sogar eine relative junge Disziplin, die Informationswissenschaft, die sich mit Informationsprozessen und Informationsproblemen in Wissenschaft, Gesellschaft, Wirtschaft und Verwaltung auseinandersetzt [Schneider 1996].

Dieses Kapitel versucht, sich dem Wesen der Information durch für Informatiker nichtalltägliche Auffassungen zu nähern. *Information, Energie und Materie*, diese drei Grundkomponenten bestimmen das Wesen unserer Welt gemäss dem Mathematiker Norbert Wiener. Dies würde bedeuten, dass Information sich mit keinem anderen Element, also auch mit keiner anderen Ware vergleichen lässt. Als eine der Konsequenzen dieser Definition von Information erscheint die Tatsache, dass Information nicht ohne Trägermedium existieren kann. Das Verhältnis von *Information und Informationsträger* reflektiert in gewissem Sinne das Verhältnis zwischen der Informationstechnologie und der Information selbst. Ohne Informationen ist jede Datenbank und jede Software, ja auch jeder Prozessor und jedes andere elektronische Gerät sinn- und nutzlos. Eine weitere Betrachtungsweise ist *der Mensch als Informationsverarbeitende Maschine*. Wir Menschen benutzen Information gemäss dieser Definition als Reduktion von Ungewissheit also um unser Wissen zu vermehren. Die Sprachwissenschaft schliesslich definiert im Gegensatz dazu Information im Rahmen der Semiotik, der allgemeinen Theorie der Zeichen, anhand der *Struktur und Bedeutung* der Information. Im Zentrum steht die Bedeutung der Symbole des primären Informationsträgers, der natürlichen Sprachen. Eine weitere interessante Auffassung ist die-

jenige der Informationswissenschaft, die *Information als handlungsrelevantes Wissen* definiert. Diese Auffassung unterscheidet zwischen Wissen und Information. Wissen kann sehr wohl redundant vorhanden sein, während eine bestimmte Information lediglich einmal vorhanden sein kann, ansonsten stellt eine Information lediglich eine Kopie bereits vorhandener Information dar. Sie ist in diesem Fall wertlos. Diese Auffassung geht noch einen Schritt weiter. Information ist erst dann wirklich Information, wenn sie zu konkreten Handlungen führt, respektive konkrete Handlungen bestimmt.

## 1.2.1 Verschiedene allgemeine Definitionen

Die wichtigsten allgemeinen Grundbegriffe für die Auseinandersetzung mit der Verarbeitung von Informationen durch abstrakte Maschinen der Informationstechnologie sind die *Informatik*, die *Kybernetik* als allgemeine Theorie der Maschinen, der *umgangssprachliche Begriff der Information* sowie die *Information aus Sicht des Informationsmanagement*.

### 1.2.1.1 Informatik

Definition Informatik des Taschenbuchs der Informatik [Schneider, Werner 2004]:

*Die Informatik ist eine Wissenschafts- und Technologiedisziplin, die sich mit Methoden und Verfahren der automatisierten Informationsverarbeitung befasst.*

### 1.2.1.2 Kybernetik

Definition Kybernetik aus dem Lexikon der Kybernetik [Laux 1980]:

*Die Kybernetik ist die Wissenschaft über die allgemeinen Gesetze der Gewinnung, Speicherung, Übertragung und Umwandlung von Information in großen Steuerungs- und Leitungssystemen. Dabei versteht man unter solchen Systemen nicht nur technische, sondern auch beliebige biologische, Verwaltungs- und gesellschaftliche Systeme.*

### 1.2.1.3 Umgangssprachlicher Begriff der Information

Neben dem umgangssprachlichen Begriff der Information als Nachricht, die einem Zweck dient, existiert die folgende Definition des Dudens:

*"In|for|ma|ti|on die; -en: 1.a) Nachricht, Auskunft; das Informieren; Unterrichtung über eine bestimmte Sache; b) [auf Anfrage erteilte] über alles Wissenswerte in Kenntnis setzende, offizielle, detaillierte Mitteilung über jemanden, etwas; c) (meist Plural) Äußerung oder Hinweis, mit dem jemand von einer [wichtigen politischen] Sache in Kenntnis gesetzt wird; d) Kurzform für Informationsstand, -stelle. 2. (Plural selten) a) [Maß für den] Informationsgehalt einer Nachricht; b) als räumliche oder zeitliche Folge physikalischer Signale, die mit bestimmten Wahrscheinlichkeiten auftreten, sich zusammensetzende Mitteilung, die beim Empfänger ein bestimmtes [Denk]verhalten bewirkt (Informatik) [...]"*

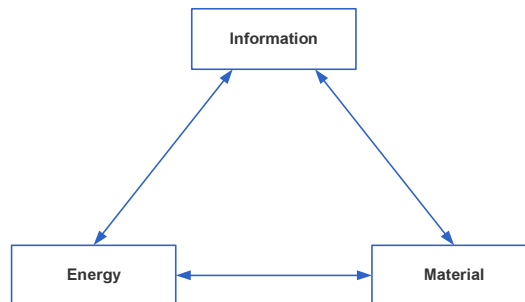
### 1.2.1.4 Information aus Sicht des Informationsmanagement

Die Definition von Information aus Sicht des Informationsmanagement ist [Heinrich, Lehner 2005]:

*„Information ist handlungsbestimmendes Wissen über historische, gegenwärtige und zukünftige Zustände der Wirklichkeit und Vorgänge in der Wirklichkeit, mit anderen Worten: Information ist Reduktion von Ungewissheit“*

## 1.3 Information als Grundkomponente der Welt

---



**Abbildung 1.2 Norbert Wieners Grundkomponenten der Welt**

Norbert Wiener, ein amerikanischer Mathematiker und Mitbegründer der Kybernetik, betrachtet Information als eine der drei Grundkomponente der Welt, die neben der Energie und der Materie existiert [Ubrig 2001]. Norbert Wiener definiert Information in seinem Buch Kybernetik: *„Information ist Information, weder Materie noch Energie. Kein Materialismus, der dies nicht berücksichtigt, kann heute überleben“* [Wiener 1982]. Nach den Vorstellungen von Wiener ist Information neben Energie und Materie eine unabhängige Grösse, die den Gesetzen von Energie und Materie nicht unterliegt. Das wichtigste Unterscheidungskriterium ist die Tatsache, dass Information weitergegeben werden kann, ohne dadurch weniger zu werden. Die Information wird durch die Weitergabe verdoppelt. Mit Materie und Energie ist der entsprechende Vorgang nicht möglich.

Information entsteht bei der Interaktion zwischen verschiedenen Komponenten eines Systems, ja sie ist für das Funktionieren des gesamten Systems, beispielsweise bestehend aus Mensch und Maschine unbedingt erforderlich. Information hat demnach aus Sicht von Wiener eine Verbesserung von Organisation und Ordnung zur Folge. Je grösser der Informationsgehalt eines Systems, desto besser ist das System organisiert. Diese Definition ist ein wesentlicher Bestandteil seiner Auffassung von Kybernetik (griechisch: Kybernetes = Steuermann) als allgemeine Theorie die Systeme beschreibt. Diese Systeme sind abstrakte deterministische Maschinen, deren Zweck es ist ein bestimmtes Ziel erreichen. Dargestellt werden diese Maschinen als Variablensysteme. In einem gewissen Sinn ist jedes Softwareprogramm ebenfalls ein Variablensystem, also auch ein kybernetisches System.

Wieners Definition von Information als drittes Grundelement unserer Welt würde auch eine Reihe von Phänomenen, die in Zusammenhang mit der Bewertung der Ware Information auftauchen, erklären. Während wir sehr wohl akzeptieren, dass Energie und Materie einen definierten Wert haben, ist der Wert von Informationen noch umstritten. Es existieren Weltmarktpreise für Energie und Rohstoffe und es existieren Warenterminbörsen für Grundnahrungsmittel. Für Information kennen wir diese Mechanismen (noch) nicht. Der Wert der Ware Information ist heute höchst umstritten. Aus Sicht von Norbert Wiener ist das dadurch erklärbar, dass sich Information grundsätzlich von Energie und Materie unterscheidet. Mehr zum Thema „Wert von Informationen“ finden Sie in Kapitel 7.2 dieser Vorlesung.

### 1.3.1 Norbert Wiener



**Abbildung 1.3** Norbert Wiener

Norbert Wiener wurde am 26.11.1884 in Columbia, Missouri geboren. Er absolvierte die High School mit 11 Jahren, das College mit 14 Jahren und machte seinen ersten Dokortitel mit 18 Jahren. Nach dem ersten Weltkrieg arbeitet er am MIT (Massachusetts Institute of Technology), um sich mit Themen wie der harmonischen Analyse, der Thermodynamik und der Quantenmechanik auseinanderzusetzen. Aus dieser Zeit entstand eine Theorie zur systematischen Minimierung von Störeffekten in Signalen, besser bekannt unter dem Begriff „Wiener Filter“. Im zweiten Weltkrieg arbeitet er an Steuerungsproblemen von Flugzeugbewegungen, die er in seinem Buch „Kybernetik: Regelung und Nachrichtenübertragung in Lebewesen und Maschine“ zusammenfasste. Norbert Wiener war ein Wissenschaftler, der regen Kontakt zu Kollegen und Kolleginnen aus anderen Fachgebieten hatte und der sich für den Einfluss der Wissenschaft auf das tägliche Leben interessiert. Ihm werden heute stehende Begriffe wie „Mensch und Maschine“, „Rückkoppelung“ und „Entropie“ zugeordnet. Wenn er auch diese Begriffe nicht erfunden hat, so hat er sich intensiv mit ihnen auseinandergesetzt und gilt heute als Begründer der Kybernetik und damit auch als Vater der Systemtheorie. Norbert Wiener wurde im Januar 1964 die "National Medal of

Science" durch den Amerikanischen Präsidenten Johnson verliehen. Er starb noch im gleichen Jahr.

Der Erfinder des Quarks, Murray Gell-Mann beschreibt Norbert Wiener in seinem Buch "Das Quark und der Jaguar" [Gell-Mann 1995]: "Als ich am MIT studierte, fand ich ihn hin und wieder schlafend im Treppenhaus liegen, wo sein fülliger Körper den Verkehr ernsthaft behinderte. Einmal steckte er seinen Kopf in das Arbeitszimmer meines Doktorvaters, Viki Weisskopf, und äusserte.... «Oh, ich dachte, alle europäischen Intellektuellen würden Chinesisch sprechen»".

## 1.4 Information und Informationsträger

---

Nach gegenwärtigem Stand des Wissens ist das Vorhandensein von Information immer an das Vorhandensein von Materie bzw. Energie gebunden. Diese dienen als Informationsträger: Information ist eine räumliche und / oder zeitliche Struktur innerhalb eines Informationsträgers, und kann daher ohne diesen nicht existieren.

### 1.4.1 Tukey und die Geschichte des Begriffs BIT



**Abbildung 1.4** John Wilder Tukey

Der Begriff Bit ist an einer Sitzung im Winter 1943/44 in Princeton festgelegt worden. Zu diesem Treffen sind Treffen von Physiologen, Rechenmaschinenkonstrukteuren und Mathematikern hatten Norbert Wiener und John von Neumann, respektive - je nach Quelle - Vannevar Bush, eingeladen. Diese Bezeichnung "Bit" wurde vom Statistiker John Wilder Tukey (1915 – 2000) [McCullagh 2003] vorgeschlagen. Information ist gemäss der Expertengruppe grundsätzlich nach der Zahl von Ja's und Neins zu messen. Diese Einheit wurde



Binary Digit oder abgekürzt Bit genannt. Tukey benutzte die Einheit dann im Zusammenhang mit der Arbeit am ENIAC-Computer 1946, was in der Folge dazu führte, dass die breite Öffentlichkeit zu der Annahme verleitet wurde, das Bit hätte ausschließlich etwas mit Computern zu tun. Zum ersten Mal in einer Publikation wurde der Begriff BIT im Juli 1948 im Artikel "The Mathematical Theory of Communication" von Claude Elwood Shannon [Shannon 1948] erwähnt.

## 1.5 Information und Mensch

---

Die Verarbeitung von Informationen durch den Menschen ist Gegenstand des Fachgebietes der kognitiven Psychologie [Eysenck, Keane 2000]. Die kognitive Psychologie versucht mit Hilfe von Modellen, menschliches Verhalten nachzubilden und damit zu formalisieren. Die Bereiche der kognitiven Psychologie umfassen unter anderem die Wahrnehmung, das Gedächtnis, die Aufmerksamkeit, die mentale Abbildung der Realität, die Problemlösungsverfahren und die Verfahren zur Entscheidungsfindung.

Die Verarbeitung von Informationen durch den Menschen ist bis heute nicht vollständig geklärt, da wir die Vorgänge im Gehirn noch nicht verstehen, sie also auch nicht in mathematischen Modellen abbilden können. Aus diesem Grund sind die Modelle der kognitiven Psychologie eines der wichtigsten Instrumente zum Verständnis der Informationsverarbeitung durch den Menschen. Oberflächlich gesehen ist die kognitive Psychologie vor allem im Bereich der Mensch-Computer-Interaktion für die Informatik relevant. Es werden dabei nach optimalen Lösungen für ergonomische User Interfaces gesucht. Betrachtet man jedoch die verschiedenen Modelle der kognitiven Psychologie oder auch der Kognitionswissenschaft genauer, so existieren dort kaum Modelle, die nicht in der Informatik angewendet werden. So folgt beispielsweise ein moderner Graphikprozessor den Prinzipien der Stufenweisen visuellen Wahrnehmung, wie sie David Marr in seinem Buch "Vision" entwickelt hat [Marr 1982]. Die "Network Theory of Semantic Memory" von Collins und Quillian aus dem Jahr 1970 wird heute durch praktisch jedes moderne Expertensystem reflektiert [Collins, Quillian 1970]. Prozessorarchitekturen wie der Pentium folgen direkt dem mehrstufigen Gedächtnismodell der Psychologen Atkinson und Shiffrin aus dem Jahr 1968 [Atkinson, Shiffrin 1968]. Dies ist nur eine Auswahl aus der Vielzahl von Erkenntnissen der Kognitionswissenschaft, die direkt in die Informatik einfließen. Mehr zum Thema Kognitive Psychologie finden Sie in Kapitel 2 dieser Vorlesung.

### 1.5.1 Der Modellbegriff

Aus der Verbindung von Kognitionswissenschaft und Kybernetik hat sich ein Konzept entwickelt, welches den Ablauf der Transformation der realen Welt in eine abstrakte und gegebenenfalls vereinfachte Darstellung dieser Welt beschreibt. Diese Darstellung wird auch das innere Modell der Welt genannt. Der Deutsche Philosoph Herbert Stachowiak hat in seinem Buch "Allgemeine Modelltheorie" bereits 1973 versucht, die Eigenschaften ei-

nes Modells zu beschreiben [Stachowiak 1973]. Seiner Ansicht nach unterscheidet sich ein Modell durch die Abbildung, die Verkürzung und den Pragmatismus gegenüber dem Original. Die *Abbildung* ist eine Repräsentation des Originals, die selbst wiederum Modelle sein können. Durch *Verkürzung* werden lediglich diejenigen Attribute des Originals im Modell übernommen, die relevant erscheinen. Der *Pragmatismus* bedeutet die Reduktion des Modells auf das Nützliche, das Modell interpretiert das Original. Das Modell abstrahiert das Original durch ein bewusstes Weglassen bestimmter Merkmale.

### 1.5.2 Das kybernetische Modell des Menschen

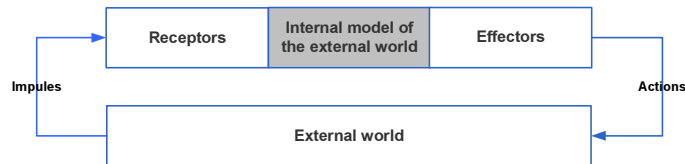


Abbildung 1.5 Mensch und Umwelt im kybernetischen Modell

- *Receptors*: Sinnesorgane des Menschen, die Informationen über Reize verarbeiten
- *Internal model of the external world*: Das zentrale Nervensystem, welches das Modell der Umwelt enthält
- *Effectors*: Der Mund und der Körper des Menschen als Organe, die die Umwelt beeinflussen können
- *External world*: Die Umwelt

Ausgangspunkt für seine allgemeine Modelltheorie war für Professor Stachowiak das kybernetische Modell des Denkens und Erkennens. In diesem Modell, welches er 4 Jahre vor seiner wichtigsten Publikation, der „Allgemeine Modelltheorie“, entwickelt hat, wird die Informationsverarbeitung zwischen Mensch und Umwelt zum Regelkreis [Stachowiak 1969]. Der Mensch besteht aus Rezeptoren, dem inneren Aussenweltmodell und Effektoren. Die Rezeptorensysteme (Augen, Ohren) empfangen die Reize aus der Umwelt. Das innere Aussenweltmodell (zentrales Nervensystem) ist eine Abbildung oder auch eine mentale Konstruktion der Aussenwelt. Handlungen, die die Aussenwelt beeinflussen, werden durch Effektorensysteme (Mund, Körper) ausgelöst. Der Mensch empfängt Reize aus der Umwelt durch Impulse und beeinflusst die Umwelt durch Handlungen. Die Umwelt wird auch Aussenwelt genannt. Informationen werden in diesem Modell durch Handlungen und durch Impulse übertragen.

### 1.5.3 Wissen und Denken

Während aus didaktischer und organisationstheoretischer Sicht der Weg von der Information zum Wissen über einen mehrstufigen Lernprozess erfolgt [Nonaka, Takeuchi 1997] definiert das kybernetische Modell des Menschen diese Prozesse wesentliche einfacher

und präziser. *Wissen* ist die Struktur des inneren Modells der Welt. *Ungewissheit* besteht, wenn für eine problematische Situation kein Lösungsweg im inneren Modell der Welt vorliegt. *Denken* sind Operationen am inneren Modell der Welt.

Daraus folgt, dass Information im weitesten Sinn die *Reduktion von Ungewissheit* ist und Information im weiteren Sinn die Reduktion von Ungewissheit durch Impulse aus der Umwelt ist. Diese Definition von Information kann nun erweitert werden. Dann ist *Information die Reduktion von Ungewissheit aufgrund von Kommunikationsprozessen (d.h. durch Zweckgerichtete Signale von anderen Organismen oder Systemen)*.

## 1.6 Information und Sprache

---

Aus Sicht der Linguistiker erfolgt der Austausch von Informationen durch den Austausch von Symbolen. Die Semiotik (oder Semiologie) ist die allgemeine Theorie der Symbole (Zeichen). Die Sprachwissenschaft (Linguistik) ein Teilgebiet der Semiotik; sie befasst sich mit den sprachlichen Symbolen. Der Genfer Sprachwissenschaftler Ferdinand de Saussure (1857-1913) hat die Semiotik zu Beginn des 20. Jahrhunderts wesentlich geprägt. De Saussure's Definition von Sprache ist: *Sprache ist ein System von Zeichen, die Ideen ausdrücken*.

Aus Sicht der Linguistik ist die Sprache dazu da, Informationen weiterzugeben. Das wichtigste Grundelement jeder Sprache ist das Symbol. Ein Symbol ist ein einzelnes Zeichen oder jedoch auch eine Abfolge von mehreren Zeichen. Information wird in einer natürlichen Sprache sowohl durch Abfolge von Symbolen also auch durch die Gruppierung solcher Abfolgen in Sätzen beeinflusst. Zur korrekten Interpretation der Symbole einer Sprache sind Syntax, Semantik und Pragmatik von Zeichen und Zeichenfolgen notwendig. Die Syntax einer Sprache umfasst das Alphabet der Sprache, also die möglichen Zeichen, die in der Sprache verwendet werden dürfen. Die Semantik der Sprache beschreibt die Bedeutung eines oder mehreren Zeichen, also die Bedeutung der Wörter und Wortfolgen. Die Pragmatik der Sprache ergänzt die Semantik mit dem Kontext des Gesprochenen. Dies unterscheidet die natürlichen Sprachen von den formalen Sprachen, wie sie in der Informatik vor allem in Form von Programmiersprachen eingesetzt werden. Programmiersprachen sind immer kontextfreie Sprachen. Nur so ist eine vollständige Interpretation der Sprache durch eine abstrakte Maschine, in diesem Fall ein Compiler, ein Browser oder ein Converter, überhaupt korrekt zu interpretieren.

### 1.6.1 Das Semiotische Dreieck der Sprache

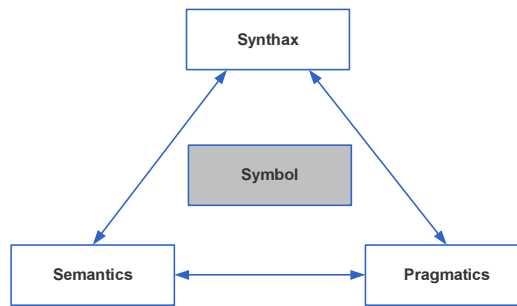


Abbildung 1.6 Das Semiotische Dreieck der Sprache

- **Syntax:** Die Schreibweise eines Symbols, d.h. alle Symbole des verwendeten Zeichensatzes.
- **Semantics:** (Griech. Zeichen bezeichnen) Die Bedeutung eines oder mehrerer Symbole.
- **Pragmatics:** (Griech. Die Kunst richtig zu handeln) Die sachbezogene Orientierung auf das Nützliche.
- **Symbol (oder Zeichen):** Zeichen sind Elemente aus einer zur Darstellung von Information vereinbarten endlichen Menge, dem Zeichenvorrat (nach DIN 44300 - Informationsverarbeitung).

### 1.6.2 Syntax

Der syntaktische Zusammenhang der Symbole umfasst die Regeln für die Verwendung der Zeichen (Abfolge, Stellung u.a.). Die Syntax einer Information beschreibt die *Auftretenswahrscheinlichkeit* eines bestimmten Symbols innerhalb eines definierten Dekodierungsschemas, die *räumliche oder zeitliche Folge* physikalischer Signale, die mit bestimmten Wahrscheinlichkeiten oder Häufigkeiten auftreten und den *Informationsgehalt* einer Nachricht, die sich aus der Anzahl der ja/nein Möglichkeiten, für die in der Nachricht einer der Werte festgelegt ist, ergibt.

### 1.6.3 Semantik

Der semantische Zusammenhang eines einzelnen Symbols oder einer Abfolge von Symbolen legt die Bedeutung der Symbole fest. Informationen werden erst verwertbar, wenn sie gelesen und interpretiert werden. Die Semantik stellt ein Bezugssystem zur Interpretation einer Information dar. Die Semantik einer Information erlaubt die Transformation, die Übertragung und das Weitergeben von Informationen. Die *Transformation* von Strukturinformation erfolgt durch Dekodierung der Symbole. Die *Übertragung* von Information als Nachricht oder Auskunft wird durch allgemeingültige semantische Regeln möglich. Das *Weitergeben* von Informationen ohne Verluste kann nur dann erfolgen, wenn durch exakte

semantische Regeln genau eine mögliche Interpretation einer bestimmten Zeichenfolge erzwingen wird.

#### 1.6.4 Pragmatik

Die pragmatische Betrachtungsweise eines Symbols berücksichtigt die Folgen und Handlungen, die mit dem Auftreten des Zeichens verbunden sind. Die Pragmatik als sachbezogene Orientierung auf das Nützliche unterstützt die Semantik einer Sprache bei der Interpretation einer Information. In der natürlichen Sprache ist die Pragmatik sehr oft ein notwendiges Mittel, um den Kontext eines Satzes zu verstehen, um eine Information richtig zu interpretieren. Der Kontext eines Satzes kann aus dem bereits Gesagtem bestehen oder jedoch durch die Betonung oder durch die begleitende Gestik gegeben sein.

Der Bereich *Natural Language Processing* benutzt Erkenntnisse aus der Linguistik, um Gesprochenes richtig zu erkennen. Diese Systeme sind heute noch wenig verbreitet, weil der Kontext oft sehr wichtig zum Verständnis einer gesprochenen Information ist. Die erfolgreichsten Systeme kombinieren die Möglichkeiten der syntaktischen und semantischen Interpretation mit adaptiven Algorithmen, die trainiert werden können. Die pragmatische Interpretation des Gesprochenen wird in diesem Fall dem User eines solchen Systems überlassen, der einem System beibringt, wie bestimmte Sätze oder Redewendungen zu interpretieren sind.

#### 1.6.5 Information, Wissen und Daten

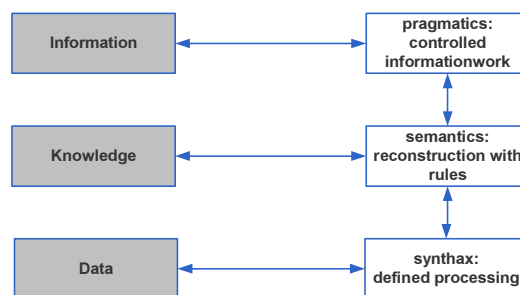


Abbildung 1.7 Semiotischer Zusammenhang Information – Wissen – Daten

- *Data*: Der Umgang mit Daten wird durch definierte Verfahren der Datenverarbeitung bestimmt. Diese Verfahren setzen eine definierte Syntax voraus.
- *Knowledge*: Die Bedeutung der einzelnen "Wissensatome", die durch die Daten beschrieben werden, ist durch die Semantik der Daten bestimmt.
- *Information*: Die Informations-Erarbeitung, -Aufbereitung und -Verarbeitung der "Wissensatome".

Der semiotische Zusammenhang zwischen Information, Wissen und Daten widerspiegelt direkt die Art und Weise, wie Informationssysteme strukturiert werden können. Die eigentlichen Daten, ob nun strukturiert in Datenbanken gehalten oder unstrukturiert in Content Management Systemen (CMS) oder Document Management Systemen (DMS) verwaltet, können lediglich auf syntaktischer Ebene verarbeitet werden. Die entsprechenden Zeichen werden einzeln gespeichert und schrittweise mittels Formatangaben oder anderen Metadaten verarbeitet. Die semantische Ebene der Verarbeitung von Daten basiert auf Regeln, die Zeichenfolgen interpretieren. Auf dieser Ebene kommen Converter, Rule Engines oder auch wissensbasierte Regelwerke zum Einsatz. Informationen können aus diesen Systemen erst dann gewonnen werden, wenn Menschen mit ihnen arbeiten. Die pragmatische Ebene umfasst sämtliche Tätigkeiten, die mittels Informationssystemen durchgeführt werden können.

### 1.6.6 Das Semiotische Dreieck der Information

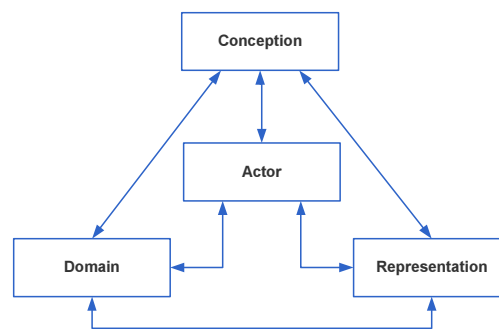


Abbildung 1.8 Das Semiotische Dreieck der Information im FRISCO Modell

- *Actor*: Diese Person ist das zentrale Element des Dreiecks. Die Person betrachtet eine Domäne, formuliert aus dieser Betrachtung heraus eine Konzeption, die sie durch eine Präsentation dargestellt.
- *Domain*: Der Betrachtungsgegenstand oder die Sachlage, die konzeptionell erfasst und in eine Form (Präsentation) gebracht wird.
- *Conception*: Die Konzeption wird durch den Actor in Zusammenarbeit oder mittels Kommunikation mit anderen Aktoren erarbeitet.
- *Representation*: Die Umsetzung der Konzeption in die physische oder symbolische Welt.

Die International Federation for Information Processing (IFIP) basiert ihr "Framework of Information System Concepts" auf dem semiotischen Dreieck der Information [IFIP 1998]. Der so genannte "FRISCO Approach" versucht die Basis für eine Theorie der Informationssysteme zu bilden. FRISCO erweitert zu diesem Zweck das semiotische Dreieck der Sprachwissenschaftler um einen Actor, der eine gegebene Sachlage konzeptionell erfasst

und darstellt. Der Report wurde durch ein Team bestehend aus europäischen und amerikanischen Wissenschaftlern und Experten erstellt. Dieses Team versuchte, Informationssysteme in ihrem organisatorischen Kontext darzustellen. Die Kommunikation von Informationen in einer Organisation mittels IT-Systemen, die Untersuchung von Meldungen und deren Bedeutung in einem sozialen Umfeld, die Ebenen der Modellierung von Informationssystemen und verschiedenen ontologischen und andere Aspekte wurden im Report untersucht [Hesse, Verrjn-Stuart 2000].

### 1.6.7 Morris und Nake's semiotische Interpretation



Abbildung 1.9 Charles William Morris

Der amerikanische Philosoph Charles William Morris (1901 - 1979) veröffentlichte 1938 sein Werk "Foundations of the Theory of Signs" [Morris 1979]. Darin formuliert er die Semiotik als Theorie der Zeichen. Ein Zeichen ist seiner Ansicht nach immer an einem Prozess beteiligt, dem Zeichenprozess oder auch der Semiose. An diesem Prozess sind der Zeichenträger (das eigentliche Zeichen), der Designat (Der Gegenstand, auf den sich das Zeichen bezieht), der Interpretant (der Effekt eines Zeichens) und der Interpret (der Empfänger des Zeichens) beteiligt. Die Semantik befasst sich gemäss Morris mit der Beziehung zwischen dem Zeichenträger und dem Designat, die Pragmatik zwischen dem Zeichenträger und dem Interpret und die Syntaktik mit derjenigen zwischen den Zeichenträgern untereinander. Der deutsche Professor Frieder Nake hat auf Basis der Zeichentheorie von Morris einen Informationsbegriff formuliert [Nake 1998]. Ein Computer kann gemäss Nake neben Rechenaufgaben allgemeine Aufgaben der Zeichenverarbeitung übernehmen. Gemäss Nake ist *Information ein Zeichenprozess, der das in einer Gemeinschaft von Kommunizierenden Unumstrittene erfasst*. Seiner Ansicht nach ist die Informatik ein Vorgang, der einen Wissensprozess algorithmisch auf einen Datenverarbeitenden Prozess abbildet.

### 1.7 Informationswissenschaft

---

Die Informationswissenschaft ist ein Begriff, der ursprünglich dazu diente, den Unterschied zwischen denjenigen Wissenschaftlern, die im Labor arbeiten und denjenigen, die sich mit der Verwaltung der wissenschaftlichen und technischen Informationen auseinandersetzen, zu definieren. *Informationswissenschaftler sollten sich mit der wissenschaftlichen Forschung des Informations- und Kommunikationsprozesses in der Wissenschaft befassen. Da das Fachwissen zur damaligen Zeit in Form von Dokumenten vorlag, entstand dafür die Bezeichnung Dokumentar (documentalist)* [Capurro 1999].

Die Informationswissenschaft untersucht Information in Zusammenhang mit Wissen. Sie versteht Information als dynamischen Prozess im Gegensatz zum statischen Wissen. Der Informationsbegriff der Informationswissenschaft geht von der Prämisse *Information ist handlungsrelevantes Wissen* aus. Demzufolge enthält redundantes Wissen keine Information.

Da die Informationswissenschaft Information als Prozess versteht, dessen Aufgabe es ist, das Wissen der Menschen zu vermehren, liegt das Hauptaugenmerk auf der Strukturierung von Informationen. Diese Strukturierung erfolgt über Techniken wie der Indexierung oder Klassierung mittels Ontologien und wird mittels Informationssystemen durchgeführt.

Die Informationswissenschaft wird vielerorts als Studiengang angeboten. Im Angebot sind Themen wie Archivsysteme, Information Retrieval, Informations-Management, Bibliothekswissenschaft zu finden.

#### 1.7.1 Definition

Informationswissenschaft an der UNI Graz :

*"Die Informationswissenschaft beschäftigt sich mit Informationsprozessen und -problemen in Wirtschaft, Wissenschaft und Gesellschaft. Sie setzt sich mit Produktion, Speicherung, Übertragung, Suche, Aufbereitung und Präsentation von Information auseinander. Darüber hinaus beschäftigt sich die Informationswissenschaft mit dem sozioökonomischen Umfeld dieser Informationstätigkeiten. Der Computer ist das wichtigste Werkzeug des Informationswissenschaftlers. Die Informations-Wissenschaft unterscheidet sich dabei insofern von der Wirtschaftsinformatik, als sie auch nicht-EDV-gestützte Systeme betrachtet."*

Information hat drei Bedeutungen aus Sicht der Informationswissenschaft [Buckland 1991]:

- *Information als Prozess:* Information ist der Prozess der Wissensmitteilung oder auch Information ist Kommunikation des Wissens.
- *Information als Wissen:* Information ist das, was im Prozess der Wissensvermittlung übermittelt wird.
- *Information als Ding:* Bestimmte Gegenstände sind informativ, weil sie die Qualität haben, Informationen zu vermitteln.



### 1.7.2 Themenauswahl

Die Informationswissenschaft befasst sich mit einer Reihe von Themen [Stock 2005]:

- Suchen und Finden von Informationen in digitalen Systemen (Websites, Internet-Portale, Datenbanken, Suchmaschinen),
- Verwalten dieser Systeme,
- Nutzergerechte Gestaltung von Informationsangeboten,
- Kreation von Methoden und Hilfsmitteln für die Abbildung von Wissen auf Informationssysteme,
- Einbeziehen des gefundenen Wissens in Arbeitsabläufe,
- Analyse von Regelmäßigkeiten bei der Ausbreitung von Informationen,
- Analyse von Informationsnutzern, deren Informationsbedürfnissen und deren Formulierungen von Anfragen an Informationssysteme,
- Unterstützung des elektronischen Handels,
- Optimieren von Informationsflüssen in Organisationen im Rahmen des Wissens-Managements,
- Markt digitaler Informationen.

### 1.8 Fragen zum Kapitel

---

- | Nr | Frage  |
|----|--|
| 1  | Wie würden Sie Information definieren?   |
| 2  | Wie viele verschiedene Definitionen des Begriffes Information kennen sie? Begründen Sie, welche der Definition Sie für richtig halten und welche Sie für falsch halten.  |
| 3  | Was finden Sie an der Theorie von Wiener interessant?  |
| 4  | Beschreiben Sie den Unterschied zwischen dem semiotischen Dreieck der Sprache und dem semiotischen Dreieck der Information.  |
| 5  | Was ist Informationswissenschaft?  |
| 6  | Studieren sie die Interaktive Studiue „Digitales Unsiversum“ von EMC und IDC ( <a href="http://www.emc.com/leadership/digital-universe/2014iview/index.htm">http://www.emc.com/leadership/digital-universe/2014iview/index.htm</a> letzter Besuch am 15.2.2015). Was halten Sie von den drei vorgeschlagenen Massnahmen? |

### 1.9 Übung zum Kapitel

---

Sie finden in den nachfolgenden Seiten den Artikel „As We May Think“ von Dr. Vannevar Bush. Herr Busch war Director des Office of Scientific Research and Development der USA. Der Artikel gilt heute als einer der ersten inhaltlichen Auseinandersetzungen zum Thema „strukturierter Zugriff auf unstrukturierte Informationen“.

Lesen Sie den beiliegenden Artikel durch und fassen Sie den Inhalt wie folgt zusammen:

1. Eine schriftliche kurze Zusammenfassung des Inhalts (max. ½ Seite A4).
2. Zeichnen Sie die Maschine, die der Artikel beschreibt, schematisch auf. Sie können mehrere Zeichnungen erstellen.