

Entropie und Information

Advanced

Wenn wir den Zusammenhang von **Entropie** und **Information** näher beleuchten wollen, müssen wir uns zunächst mal anschauen, wie man Information *quantitativ* definiert.

- Die erste wichtige Erkenntnis dazu ist: Das geht überhaupt nicht so, wie man sich das naiv vorstellen würde. Schauen wir ein Beispiel an:
- Welcher der zwei nachfolgenden Symbolketten enthält mehr Information?

1. **PRÜFUNG BESTANDEN**
2. **ARFGUND SEPÜNNBET**

Die an sich klare Antwort "1" ist falsch.

- Denn die beiden Sätze enthalten technisch, d.h. von der **Syntax** (= Muster, Zusammenstellung) her gesehen, exakt dieselbe Information, denn die beiden Zeichenketten bestehen aus einem identischen Satz von Zeichen.
- Die Zahl an *Bits*, die man braucht, um beide Sätze zu kodieren und dann vielleicht auf irgendeinem Datenkanal zu übertragen ist identisch. Auch die Störanfälligkeit, der mögliche Verlust an Information ist dieselbe für die beiden Sätze sowie für alle anderen möglichen Sequenzen, die sich aus der gegebenen Menge an Symbolen bilden lassen.

Dass *wir* das Gefühl haben, dass eigentlich nur eine einzige Sequenz (die Nr. 1) Sinn ergibt und damit Information enthält, liegt daran, dass *wir* automatisch die **Semantik** (= Bedeutung) einer Symbolsequenz betrachten.

- Maschinen können das aber nicht; siehe das grandiose Scheitern der "**künstlichen Intelligenz**".
- Es gibt deshalb keine befriedigende Definition von Information, die auch der *Bedeutung* der betrachteten Information gerecht wird; selbst nur auf die Syntax bezogene Definitionen sind im Grunde noch unbefriedigend.

Da wir aber nichts Besseres haben, nehmen wir die klassische Definition, von **Shannon 1948** eingeführt.

Betrachten wir zunächst eine Symbolmenge, z.B. ein Alphabet, mit **N** Zeichen oder Symbolen. Damit treten in *statistisch* gebildeten Symbolketten alle Symbole gleich häufig auf (im Gegensatz zu *Wörtern* einer Sprache, z.B. der Deutschen, in denen z.B. das Symbol "E" sehr viel häufiger auftreten wird als "Y").

- Damit ist die Wahrscheinlichkeit p_i dafür, dass ein Symbol vorkommt, für alle Symbole gleich groß und wir haben $p_i = p = 1/N$
- Nebenbei bemerkt: Wenn wir die p_i unterschiedlich groß machen würden und dann anfangen Symbolketten zu bilden, landen wir ganz schnell bei den Prinzipien der *statistischen Thermodynamik*.

Wieviel Information steckt in *einem* Zeichen?

- Stellen wir uns vor, wir warten bei der Übermittlung einer Sequenz auf das nächste Symbol. Wir definieren die im Zeichen steckende Information *I* über die Formel

$$N = 2^I$$

- I* ist dann schlicht die Zahl der Ja/Nein Fragen, die man braucht um mit der *geringstmöglichen* Anzahl an Fragen herauszufinden, um welches Symbol es sich handelt
- Das "geringstmöglich" ist dabei ein bißchen unpräzise. Wir würden aber z.B. bei einem Alphabet uns nicht naiv wie folgt durchfragen: Ist es **A**? - Nein, Ist es **B**?, ...; sondern fragen: Kommt es aus der 1. Hälfte? - Nein; 1. Hälfte der 2. Hälfte?,... . Nach jeder derartigen Frage bleibt dann immer nur noch die Hälfte der Möglichkeiten.

Damit haben wir eine 1. Definition für Information

$$I = \text{ld } N = \text{ld } (1/p)$$

- Dabei steht "**ld**" für den "Logarithmus dualis", den Logarithmus zur Basis 2.

Das ist aber zu einfach um damit viel anfangen zu können, Deshalb unterstellen wir jetzt, dass die Einzelwahrscheinlichkeiten p_i des Auftretens der Symbole verschieden sind - so wie in einem "richtigen" Alphabet im Kontext einer "richtigen" Sprache.

- Die Information $I(\mathbf{z}_i)$, die im *i*-ten Symbol (= \mathbf{z}_i) steckt, ist dann offenbar

$$I(\mathbf{z}_i) = \text{ld } N = \text{ld } (1/p_i) = -\text{ld } (p_i)$$

- Unwahrscheinliche Buchstaben enthalten dann mehr Information als wahrscheinliche. Das ist auch OK, denn wenn wir in einer Übertragung ein "y" oder "x" bekommen, wissen wir einfach mehr als wenn mal wieder ein "e" überwächst.

So gesehen können wir den Informationsgehalt einer Nachricht auch wie folgt definieren:

Der Informationsgehalt eines übermittelten Symbols ist proportional zum Grad unserer **Überraschung**.

● Wenn wieder mal ein "e" kommt, hält sich die Überraschung in Grenzen, aber ein "y" erwarten wir halt eher nicht¹⁾.

Im nächsten Schritt interessieren wir uns nur noch für den **Mittelwert H** der Information, der in einem durch die p_i definierten spezifischen sprachbezogenen Alphabet mit N Symbolen steckt.

● Um einen zum Mittelwert proportionalen Wert H zu erhalten, müssen wir über die in den Symbolen enthaltene Information **multipliziert mit der Wahrscheinlichkeit ihres Auftretens** summieren; wir haben also

$$H = \sum_{i=1}^N p_i \cdot I(p_i) = - \sum_{i=1}^N p_i \cdot \ln(p_i)$$

Das ist die klassische Formel von Shannon. Die Größe H hat er, und das sollte uns jetzt nicht all zu sehr verblüffen, **Entropie** genannt.

● Die Bedeutung von H für die Informationstheorie liegt darin, dass wenn wir Z Zeichen übertragen wollen, das Produkt $H \cdot Z$ direkt die Mindestzahl der **bits** angibt, die man braucht um mit dem Alphabet Information übertragen zu können.

Aber warum nennt Shannon diesen Informationsmittelwert **Entropie**? Ist das tatsächlich dasselbe, wie die Entropie in der Thermodynamik?

Nun ja - nicht exakt. But close enough. Von der thermodynamischen Entropie S , wie sie in der [statistischen Thermodynamik](#) definiert wird, unterscheidet sich H nur in zwei verhältnismäßig trivialen Punkten:

● 1. Die oben indirekt angesprochene **Proportionalitätskonstante** (die in den Gleichungen immer = 1 gesetzt ist) muss $k =$ Boltzmannkonstante sein. Aber das ist wahrlich trivial; wir messen nur mit verschiedenen Maßsystemen.

● 2. Die korrekte thermodynamische Entropie, die eigentlich nur für Gleichgewicht definiert ist, entspricht genau genommen nur dem **Maximalwert** von H , den wir für Gleichverteilung der Symbole erhalten.

Wenn man dann das alles gebührend berücksichtigt, erhält man eine interessante Beziehung für die **thermodynamische Entropie**, die in einem bit Information steckt:

$$S(1 \text{ bit}) = -k \cdot \ln 2$$

● Das bedeutet, dass eine Entropieerhöhung von $0,957 \cdot 10^{-23} \text{ JK}^{-1}$ in einem gegebenen System, genau ein **bit** an Information vernichtet. Und da die Entropie in einem abgeschlossenen System nicht abnehmen kann, kann auch keine Information spontan entstehen.

● Das ist nicht nur "Theorie", sondern damit konnte Leo **Szilard** zum ersten Mal eines der berühmtesten offenen Probleme der Thermodynamik, den "Maxwellschen Dämon" erklären. Mal selbst Googeln; Stichworte: information entropie maxwell dämon.

Immerhin, irgendwie scheinen die gute alte **Entropie** und die noch recht junge und nicht sonderlich gut definierte **Information** zusammenzuhängen. Hat das was zu bedeuten? Wird die Informatik möglicherweise auch mal zu einem Untergebiet der Physik, so wie die Chemie?

● Nichts genaues weiß man nicht. Es gibt berühmte Leute, die sehr konträre Ansichten vertreten. [Atkins](#), z.B. hält das alles für trivial bis Blödsinn, während [Penrose](#) darin einen der Schlüssel für die künftige "neue" Physik sieht.

● Wir werden sehen.

¹⁾ Das häufigste Zeichen der deutschen Sprache ist übrigens das " ", der Leerschritt, mit einer Wahrscheinlichkeit von 0.151. Danach kommen E (0.147), N (0.088), R (0.068) und I (0.063). Schlußlichter sind Y (0.000173), Q (0.000142) und X (0.000129).