

## Der Anteil an Wissen in Bibliotheken

### 1. *Einleitung*

Nach der Entwicklung der Informationstheorie in der ersten Hälfte des 20. Jahrhunderts<sup>1</sup>, in der man entdeckt hatte, dass Information in Bit messbar ist, begann man mit der Abschätzung des Informationsgehaltes in einem Buchstaben, einem Wort, einem Satz, einem Buch und 1963, im Weinbergreport, auch mit dem der Library of Congress (LC) in den USA. Der praktische Hintergrund war, dass man schon damals überlegte, die gesamte Information der LC digital zu speichern. Da das nicht möglich schien, und damals auch utopisch war, entschloss man sich zunächst mit den Bibliografien, ERIC, INDEX MEDICUS, BIOLOGICAL ABSTRACTS, CHEMABS, CAIN und dem Katalog der LC zu beginnen, zum Nachweis dessen was an wichtigen Publikationen überhaupt vorhanden ist. Dies war im Prinzip der Startschuss für die Digitale Bibliothek, in der zunächst die Bibliographien und Kataloge digitalisiert wurden. Später kamen die Abstracts, die Volltexte und am Schluss auch die Bilder, Grafiken und Tabellen hinzu. Heute werden nicht nur gedruckte Dokumente, wie Bücher, Patente oder Zeitschriften digital angeboten, sondern auch Tondokumente, Filme und alle anderen multimedialen Publikationen.

Die grobe Abschätzung von 1013 Bit Information in der LC, erfolgte 1963 in sehr vereinfachender Weise. In dem man etwa sieben Bit pro Buchstaben, zwei bis drei Tausend Zeichen pro Seite, etwa zweihundert Seiten und rund fünf Millionen Bücher zusammenrechnet, kommt man auf diesen Wert. Diese Abschätzung implizierte aber damals drei grundsätzliche Fehler, auf die wir inzwischen aufmerksam wurden, weil unsere heutigen Computer weitaus mehr, qualitativ höher und auch intelligenter Bücher in digitalisierter Form verfügbar zu machen vermögen. Das Wort intelligent bezieht sich dabei auf die Fähigkeit der Rechner, aus vorhandenen Daten die wirklich überflüssige Redundanz bedarfsgerecht zu eliminieren.

1 Shannon, C.E., u. Weaver, W., The mathematical theory of communication. Urbana: University of Illinois Press. (1949).

Die drei Fehler waren:

1. Die Bilder wurden vollständig vernachlässigt.
2. Die Redundanzen wurden vollständig vernachlässigt,
3. Die Tatsache, dass das Maß zur Bestimmung von Information, Redundanz, Rauschen und Wissen, das Bit, keine lineare Skalierung aufweist, wie wir sie beim MKS-System kennen, wurde vernachlässigt.

Während sich die Fehler 1. und 2. in erster Näherung gegenseitig weitgehend kompensieren, wirkt sich der dritte Fehler auf unsere Einschätzung des Anteils an Information und insbesondere an Wissen in Bibliotheken, und damit auch in dieser Welt, geradezu fatal aus.

- Zu Punkt 1.: Die digitale Speicherung von Bildern verbraucht, je nach Raster und Farbtiefe, etwa das Tausendfache an Speicherkapazität dessen, was eine reine Textseite erfordert. Geht man von einer Größenordnung von 10% Bildmaterial in Büchern aus, so hätte der Informationsgehalt der LC mit dem Hundertfachen, also 1015 Bit angegeben werden müssen.
- Zu Punkt 2.: Kompressionsprogramme können ohne nennenswerten Informationsverlust, dieses Aufkommen von 1015 Bit an Daten, weil sie nicht nur Information, sondern in hohem Maße auch Redundanz und Rauschen enthalten, um den Faktor Hundert wieder auf 1013 Bit reduzieren.
- Zu Punkt 3.: Der Fehler, das Maß für Information und Wissen meist linear zu berechnen, ist weit verbreitet und lässt sich oft auch nicht ganz vermeiden, es ist aber wichtig, auf den gedanklichen Bruch hinzuweisen, wenn wir beispielsweise 256 ASCII-Zeichen mit 28 Bit codieren, aber bei einem Wort wie „Byte“  $4 * 28$  Bit, also linear, weiterrechnen.<sup>2</sup>

Beim Wissen handelt es sich, wenn man die Erkenntnis darüber konsequent auf die Informationstheorie stützt, um begründete Information bzw. evidence based information. Diese Erkenntnis, die damit eine klare Definition dessen was Wissen ist einschließt, ist im Grundsatz nicht neu, sie geht ebenso auf die griechische Philosophie, wie auch auf Überlegungen zur Zeit der Aufklärung zurück, sie erfährt aber eine Präzisierung durch die Informationstheorie, da erst diese einen klar messbaren Informationsbegriff erbrachte.

2 Umstätter, W.: Die Skalierung von Information, Wissen und Literatur. Nachr. f. Dok. 43(1992)(4), S. 227 – 242.

## 2. *Biologische Wissensspeicher*

Im Vergleich zu den bibliothekarischen Speichern, die stark durch die linearen Buchstabenketten geprägt sind, hat man bei Chromosomen bis zu 2 Meter lange DNS-Stränge (Desoxyribonukleinsäure) gefunden. Sie enthalten in ihrer Makromolekülkette Abwechslungen und damit Codierungen, die etwa einem Bit pro Nanometer entsprechen, woraus sich etwa  $2 \times 10^9$  Bit an Information, Redundanz und Rauschen bei einem maximalen Chromosom ergeben. Das entspricht etwa 250 MB, oder einer halb gefüllten CD-ROM. Bei einer Buchstabenfolge in einem Buch haben wir anstatt des einen Bits/Nm etwa 1 Bit/mm, also etwa ein Millionstel der Informationsdichte der Biologie.

Die Erkenntnis, dass die DNS ein Molekül ist, das Erbinformation in dieser Codierung trägt, ist eng mit der Informationstheorie verbunden, da erst diese Theorie die Zusammenhänge erkennbar machte. Vorher spekulierte man noch über geheime Kräfte und Stoffe, die dem Leben innewohnen müssten und denen man Namen wie Entelechie oder Elan Vital gab.

Insgesamt soll das menschliche Genom etwa 3,3 Mrd. Basenpaare enthalten. Da bei genauer Betrachtung mit jeweils einem Tripel solcher Basenpaare 20 Aminosäuren codiert werden, kommen 4,3 Bit of ein Codon bzw. 1,4 Bit pro Nukleotid. Das entspricht  $4,6 \times 10^9$  Bit. Das sind bei den 23 menschlichen Chromosomen durchschnittlich  $2 \times 10^8$  Bit pro Chromosom bzw. 20 cm lange DNS-Stränge. In jeder der etwa 10<sup>12</sup> Körperzellen befindet sich diese Speicherkapazität in doppelter Menge. Männer haben durch ihr y-Chromosom etwa 2% weniger gespeicherte Erbinformation.

Obwohl wir hinsichtlich des sog. Processings recht brauchbare Kenntnis darüber besitzen, wie diese Information der DNS in den einzelnen Körperzellen sozusagen parallelprozessierend, über Transcription, Translation abgelesen und zu Enzymen bzw. Strukturproteinen weiterverarbeitet wird, fehlt uns trotzdem noch immer eine weitergehende Einsicht darin, wie viel Redundanz und Rauschen in diesem System mit integriert ist, so dass wir auch hier die Skalierung der Information korrekt kalkulieren können. Ein Teil ist zweifellos exponentiell skaliert, ein anderer Teil eher linear.

Aus der Menge an biologisch nuklearer Speicherkapazität, so wie sie bei den evolutionär verschieden hohen Lebewesen vorliegt, kann man aber weitere Einsichten gewinnen. So vererben einzellige Amöben etwa  $4 \times 10^8$  Bit, während das Bakterium *Escherichia coli* mit 5,7 Mio. Basenpaaren  $8 \times 10^6$  Bit, und damit nur 1,7 Promille dessen, was der Mensch vererbt, an Erbgut überträgt. Ein Virus kann dagegen schon mit 10<sup>4</sup> Bit existieren, da es sozusagen als geborgtes Leben

existiert. Es nutzt die Zellausstattung und damit den Stoffwechsel anderer Lebewesen, in denen es sich vermehrt.

Fünfzig Prozent des genetischen Materials hat als hoch repetitiver Anteil, wie beispielsweise ...ATATATATATAT... für uns heute noch keine erkennbare Funktion. Es beherbergt also eine sehr hohe Menge an Redundanz, so dass wir nur schwerlich von Erb-Information sprechen können.

Der Mensch hat rund 45.000 Gene mit durchschnittlich 2.000 Genen pro Chromosom. Diese Unterteilung des Genoms in Chromosomen sagt etwas über die Mischung des Genmaterials der Eltern bei der Vermehrung aus. Ihre Zahl beträgt bei Menschen 23, bei Pferden 64, bei Hunden 78 und bei Farnen 512. Wir sehen damit, dass bei den ursprünglichen Farnen sehr viel mehr Kombinationsmöglichkeiten bestehen, als beispielsweise beim Menschen. Hier können aber durch das sogenannte crossing over oder durch andere Chromosomenmutationen auch durchaus weitere Genkombinationen vorkommen, und damit von der Natur getestet werden. Im übertragenen Sinne könnte dies mit den Theorien in einer Bibliothek verglichen werden, die mit anderen Theorie in Verbindung gebracht werden, und sich damit als widersprüchlich oder auch als ergänzend erweisen können.

Da die Gene, entsprechend dem Biogenetischen Grundgesetz, noch einen großen Teil ihrer Phylogenie erkennen lassen, kann man aus ihren Verwandtschaftsgraden heraus auf ihre gemeinsamen Wurzeln zurückrechnen. Demnach wird vermutet, und stochastisch zurückgerechnet, dass alle Menschen von einem gemeinsamen Individuum abstammen, das vor ~270.000 Jahren lebte. Entsprechend liegt der Vorgänger von Mensch und Affe 7 Mio. Jahre zurück, während der gemeinsame Vorfahre von Mensch und Maus vor 50 Mio. Jahren gelebt haben müsste. Die jeweilige Rückrechnung auf ein gemeinsames Individuum dürfte allerdings eine Vereinfachung des Problems beinhalten, da wir zumindest bei einigen Rassen die Erfahrung gemacht haben, wie z.B. bei den Hunderasse Dingo bzw. Dobermann, oder bei den Kladruber Pferden, dass gerade die Mischung von Chromosomen zu charakteristischen phänotypischen Ausprägungen führten, die sich dann stabil weiter vererbten. Ob allerdings auch Arten so entstanden sein könnten bleibt unklar.

Das Auftreten der ersten Zelle wird heute auf etwa 3,5 Mrd. Jahre zurückgerechnet, also etwa 1 Mrd. Jahre nach der Entstehung der Erde, die allerdings erst vor 3,7 Mrd. Jahren überhaupt von Leben (Bakterien) bevölkert werden konnte. Der Urknall liegt allerdings schätzungsweise 15 Mrd. Jahre zurück.

Die hohe Redundanz, im genetischen Material zeigt sich nicht nur in der Zahl der bereits erwähnten somatischen Zellen und der Zahl an Individuen, beim Menschen sind das weitere  $6 \times 10^9$ , sie wird auch in der Polyploidie sichtbar.

Sieht man von den zahlreichen Mutationen ab, die in diesem gesamten Genmaterial vorhanden sind, so ergibt sich die fast unvorstellbare Menge von rund 1022 quasi Duplikaten. Auch wenn diese sicher nicht identisch sind, so stimmen sie in hohem Maße überein.

Es ist unwahrscheinlich, dass der hohe Anteil an Redundanz überflüssig ist. Sehr viel wahrscheinlicher ist er zur Erhaltung der Arten einerseits und zur evolutionären Anpassung andererseits notwendig.

Dagegen ist die vergleichbare Redundanz der publizierten Information völlig anders verteilt und unvergleichlich geringer. Ein hoher Anteil dessen ist nur in etwa 100 – 102-facher Kopie verfügbar. Insofern ist es nicht verwunderlich, dass etliche Informationen im Laufe der Geschichte weitgehend zerstört worden sind, einfach verloren gingen oder die nur noch von ganz vereinzelt Archiven oder Bibliotheken aufbewahrt wurden. Andere Bereiche an Information verteilen sich auf zahlreiche der schätzungsweise halben Million an Bibliotheken der Welt, auf eine unüberschaubare Zahl an Privathaushalten und natürlich auch auf die Erinnerung der Menschen selbst. Bezogen auf die Bibel erschien die erste gedruckte Ausgabe mit beweglichen Lettern schätzungsweise in 180 Exemplaren 140 Papier- und 40 Pergamentexemplare). Zwischen 1624 und 1665 erschienen schon 140 als 15 Ausgaben, mit Auflagenzahlen zwischen 3.000 und 6.000.<sup>3</sup>

Nach dem Bericht der Vereinigten Bibelgesellschaften von 1998 (Scripture Distribution Report), hatten allein die ihnen angeschlossenen Gesellschaften 20,8 Millionen vollständige Bibeln und 20,1 Millionen Testamente verteilt. Damit scheint eine Verbreitung von Bibeln, mit rund 80.000 Exemplaren um 1800, schon 1,5 Mio. im Jahre 1846 und bereits 150 Mio. 1900 durchaus wahrscheinlich. Das entspricht einem grob geschätzten Wachstum von über 7% jährlich, wobei sicher immer wieder einige Prozent verloren gingen. Dagegen machen die geschätzten Zuwachszahlen von 21 Mio. im Jahr 1950 und etwa 150 Mio. im Jahr 2000 ein deutlich verlangsamtes Wachstum erkennbar, was allerdings bei einer Erdbevölkerung von 6 Mrd. Menschen nicht verwundern kann. Im Gegenteil, demnach würde heute auf jeden 40sten Menschen pro Jahr eine neue Bibel kommen, und wenn diese Zahl stimmt, würde jeder zweite Mensch der zur Welt kommt, rein rechnerisch eine Bibel erhalten. Die Bibel liegt also durchaus im Bereich von 109 hoch redundant vor.

Der Anteil publizierten Wissens ist allerdings weitaus redundanter, als der der unbegründeten Informationen, da ein und dasselbe Wissen an zahllosen Stellen

3 Ausstellung „Heilige Schriften Bibeln und religiöse Texte aus 1000 Jahren“ in der Paulinerkirche, Historisches Gebäude der SUB Göttingen vom 9. März bis 27. April 2003  
<http://www.paulinerkirche-goettingen.de/bibeln.pdf>

wieder und wieder gebraucht wird und sich auch selbstreproduzierend erzeugt. Dagegen können Informationen beliebig viele Varianten mit zahllosen Details enthalten.

Das allgemein vorhandene Wissen in den Köpfen von Menschen stimmt über weite Bereiche überein und erreicht in diesen Fällen ebenfalls Redundanzen von etwa 109. Es sei nur an all die Fähigkeiten erinnert, wie man atmet, geht, liest, isst, sich kämmt, schwimmt, spricht oder wäscht. Abgesehen von den vielen kleinen persönlichen Nuancen ist das unbewusst phylogenetisch und unbewusst gespeicherte Wissen für all diese Handlungen weitgehend identisch. Man schätzt die Kapazität unseres Gehirns heute auf 10 Mrd. Neuronen und rund 1.000 Synapsen pro Neuron. Daraus ergibt sich größenordnungsmäßig eine Kapazität von 1013 Bit.

Einige Nervenzellen in der Hirnrinde dürften sogar mit etwa 200.000 anderen kommunizieren. Wobei das Gehirn der Säugetiere eher mit deren Körperoberfläche als mit deren Volumen korrespondiert, was vermuten lässt, dass die Neuronen des Gehirns weitaus mehr mit den Informationen aus der Umwelt, als mit denen des Körpers selbst beschäftigt sind. Für den vegetativen Bereich ist ohnehin mehr das Rückenmark zuständig. Außerdem handelt jede Zelle, da sie ja mit der gesamten Information des Organismus ausgestattet ist, weitgehend selbständig.

Unser Kurzzeitgedächtnis, das etwa 15 Sec. reicht, verarbeitet etwa 7 Bit. Danach werden die Inhalte vergessen, oder sobald sie eine erhöhte Bedeutung für uns haben, in das Langzeitgedächtnis transportiert. Das Kurzzeitgedächtnis fungiert somit als wichtigstes Filter beim Import in das Langzeitgedächtnis und in unser Bewusstsein. Aus bibliothekarischer Sicht gehören gesellschaftspolitisch betrachtet alle Modeerscheinungen, die nicht in Bibliotheken, Archiven oder Museen gespeichert sind, sozusagen zum Kurzzeitgedächtnis der Menschheit.

Einen weiteren Filtermechanismus, wie beim Übergang vom Kurz- zum Langzeitgedächtnis, können wir beim sehen beobachten. Die Augen verarbeiten mit Hilfe ihrer Zäpfchen und Stäbchen sozusagen parallelprozessierend ca.  $2 \times 10^8$  Bit/s. Die ableitenden Nervenfasern aber nur noch  $2 \times 10^6$  Bit/s, und in unser Bewusstsein gelangen davon nur noch etwa 20 bis 200 Bit/s. Diese radikale Abstraktion ist möglich, weil es beispielsweise bestimmte Synapsen gibt, die für horizontale, für vertikale, für schräge, für bewegte Linien etc. zuständig sind. Sie klassifizieren für uns unbewusst unsere gesamte Umwelt und abstrahieren sie damit drastisch. Dabei erkennen wir Strukturen, aus der Entdeckung solcher Redundanzen heraus.

Die Abstraktion des zu sehenden, zu hörenden oder auch zu fühlenden basiert auf Jahrmillionen biogenetischer Erfahrung und stellt, so wie jede Klassierung,

informationstheoretisch eine Redundanzierung (eine Umwandlung von Information in Redundanz durch die Schaffung von Äquivalenzklassen) dar.

So sehen wir bei einem grünen Rasen nicht alle Grashalme einzeln, sondern gewinnen diesen Eindruck nur dadurch, dass wir im Zentrum unseres Auges solche Grashalme im Detail erkennen, und sie dann, durch das nur noch unscharf erscheinende Umfeld visuell extrapolieren. Das Auge sucht informationstheoretisch somit gezielt nach einer Information und ihrer Redundanz. Dass diese Mischung in unserem Gehirn, also in unserer Vorstellung, den Eindruck eines kompletten Bildes hervorruft, zeigt, dass wir auf diesem Wege, mit erstaunlich wenig Bit an Information und Redundanz, ein inneres Modell von dieser Welt erzeugen können. Der größte Teil dessen erfolgt unbewusst und aus phylogenetischer Erfahrung heraus.

Im Grundsatz ist dieser Vorgang beim sehen, hören, riechen, schmecken oder fühlen, bei Mensch und Tier zweifellos identisch. Er unterscheidet sich lediglich darin, welche Details, bzw. Charakteristika herausgefiltert und dann extrapoliert werden. Tiere entwickeln somit eigene innere Modelle von dieser Welt, denn wir müssen davon ausgehen, dass ein Regenwurm eine eigene Merk- und Wirkwelt hat, wie es J. von Uexküll nannte.<sup>4</sup>

Wenn es Wissen in dieser Welt gibt, so nehmen wir es am sichersten in den Köpfen von Menschen an. Dabei müssen wir zwei Arten des Wissens unterscheiden, das Wissen das wir im Laufe unseres Lebens erwerben und das, was wir schon ererbt haben. Die zweite Art von Wissen, die uns zunächst unbewusst ist, hat die Wissenschaft im Laufe der Zeit über das Unterbewusstsein immer deutlicher zutage gefördert. Dazu gehört die große Zahl an Mustererkennungen, die Erkenntnis von Gesichtern, die des Kindchenschemas, die von gefährlichen Geräuschen, von abstoßenden oder attraktiven Gerüchen, etc. Im allgemeinen wird dieses Wissen vom menschlichen Bewusstsein nur sehr partiell kontrolliert, so dass wir davon ausgehen müssen, dass der unbewusste Anteil weit größer ist, als der, den wir uns heute bewusst machen können. Dieses ererbte Wissen haben wir über die Biogenetische Evolutionsstrategie über Jahrmillionen der Phylogenie erworben.<sup>5</sup> Damit ist auch klar, dass diese Art des Wissens schon in Pflanzen und Tieren vorkommt, und deren Überleben seit Jahrmillionen gesichert hat. Auch K. Popper kam daher zu der Erkenntnis: „Also die Blumen »wissen« etwas über allgemeine Regelmäßigkeiten.“<sup>6</sup>

4 Uexküll, J. von, Kompositionslehre der Natur. Ullstein Verlag 1980. S. 219.

5 Umstätter, W., Die evolutionsstrategische Entstehung von Wissen. Fortschritte in der Wissensorganisation Band 2 (FW-2), Hrsg. Deutsche Sektion der Internationalen Gesellschaft für Wissensorganisation e.V. S.1 – 11, Indeks Verlag (1992).

Dagegen kam das Bewusstsein sehr spät in die Evolution des Lebens hinein. Soweit wir es heute beurteilen können, erwarb erst der Mensch diese Fähigkeit, Wissen über sein Wissen zu erwerben. Er kann sein unterbewusstes Wissen analysieren, kann es in den ihm bekannten Bereichen kontrollieren, und damit auch komplexere Probleme lösen, die von Tieren beispielsweise nicht gelöst werden können. Es sollte aber nicht unterschätzt werden, wie komplex das Wissen von Tieren bereits ist, das in deren innerem Modell existiert. Dazu gehören u.a. bereits vorbereitete Lernvorgänge, wie Prägungsphasen, angepasste Jagdstrategien, Flugserfahrungen bei Vögeln und vieles mehr. Im Gegensatz zur landläufigen Meinung, dass Vögel erst fliegen lernen müssen, bevor sie sich in das Reich der Lüfte erheben, haben sie das Wissen zu dieser Fähigkeit ebenso wie die Fledermäuse, Schmetterlinge oder Libellen ererbt. Der fälschliche Glaube, die Jungvögel müssten erst üben, weil man sie wiederholt schwingenschlagend am Nestrand sieht, entsteht lediglich durch die Tatsache, dass diese Tiere die Neigung fliegen zu wollen früher verspüren, als es die anatomische und physiologische Entwicklung bedingt. Diese Neigungsstruktur im Verhalten, die ja immer mit den anderen Veranlagungen synchron verlaufen sollte, scheint nicht selten etwas vorauszuweichen. So beobachtet man auch bei Kindern nicht selten, dass sie laufen oder sprechen wollen, bevor diese Veranlagungen schon voll ausgebildet sind.

Nimmt man diese Wissensarten zusammen, die wir einerseits als tacit knowledge und andererseits als Handlungswissen klassieren, so beinhalten sie zweifellos weit mehr gespeichertes Wissen, als das, was uns bewusst ist, und was wir von dem uns bewussten Wissen in Bibliotheken schon deponiert haben. Wir haben so betrachtet in den größenordnungsmäßig 10<sup>9</sup> Bit genetisch gespeicherter Erfahrung, aller Wahrscheinlichkeit nach noch erhebliche Mengen an Redundanz, aber auch an Rauschen. Wenn darin trotzdem mehr Wissen gespeichert sein dürfte, als beispielsweise in der Library of Congress, so liegt das natürlich zunächst an der weit aus kompakteren Speicherform dieses Wissens im inneren Modell. Zum anderen lässt es uns aber ahnen, dass vermutlich von den 10<sup>13</sup> Bit in der LC nur etwa ein tausendstel oder weniger wirkliches Wissen ist. Hier gewinnen wir einen kleinen Einblick in die Tatsache, dass Information und Wissen einer exponentiellen Skalierung unterliegen. Während bei 10<sup>1</sup> Bits über eine dreistellige Zahl entschieden wird, ist es bei 10<sup>2</sup> Bits schon eine dreißigstellige Zahl. Wir haben es also auch in den Bibliotheken mit geradezu unvorstellbaren Mengen an Redundanz und Rauschen zu tun, so dass unsere lineare Einschätzung der Informationsmenge eher den Tatsachen entspricht, als die theoretisch korrektere exponentielle.

6 Popper, K., Alles Leben ist Problemlösen. Über Erkenntnisse, Geschichte und Politik. S.135 München: Piper Verlag 1994.

Unter dem informationstheoretischen Rauschen im genetischen Bereich müssen wir in diesem Zusammenhang den Anteil einstufen, der z.B. mutierte Gene enthält, die damit aber nicht wirksam werden, so wie wir auch in Bibliotheken unleserliche Texte, unscharfe Bilder oder auch undefinierbar graue Papierflächen finden können. Dieses Rauschen kann aber durchaus zu interessanten oder wichtigen Informationen mutieren, wenn wir z.B. in alten Handschriften ungesehene wiederentdecken.

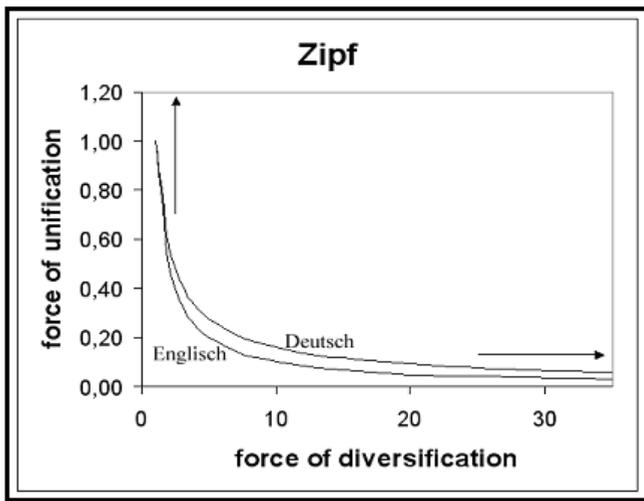
Um korrekt zu sein, muss hier allerdings auch darauf hingewiesen werden, dass das Wissen in Bibliotheken nur begrenzt mit dem Wissen einer Person vergleichbar ist. Während unser Körper genau weiß, wie er die vielfältige Nahrung, die wir zu uns nehmen, aufspaltet, in Energie umsetzt und ausscheidet, erklärt uns die publizierte Biochemie nur geringe Bruchteile dieser komplexen Abläufe. Außerdem ist, wie wir bereits sahen, das Wissen, das im Chromosomensatz eines Menschen gespeichert ist, nicht identisch, mit dem, das dann in der Ontogenie, Zelle für Zelle frei gesetzt wird. Allein im Gehirn werden daraus etwa 1012 Bit Speicherkapazität entfaltet. Die Kompression der als Wissen in der DNS gespeicherten Information ist daher beeindruckend hoch.

Merkwürdigerweise wird oft Einstein mit der Aussage zitiert, dass der Mensch vermutlich nur 10 Prozent seines Gehirns wirklich zum Einsatz bringt. Es gibt aber eine Reihe von Indizien dafür, dass diese Einschätzung in der Größenordnung stimmen könnte, da das Gehirn bei Verletzungen, Alterung oder Krankheit die entsprechenden Verluste oft durch Aktivierung anderer Gehirnbereiche zu kompensieren in der Lage ist. Wenn wir somit statt von 1012 Bit von „nur“ 1011 Bit genutzter Kapazität ausgehen, dürfte dies durchaus realistisch sein. Gehen wir weiterhin davon aus, dass 90% der Information in unserem Gehirn keinesfalls begründet, sondern nur stupide auswendig gelernt ist, kommen wir wieder in den Bereich von etwa 1010 Bit, an Wissen.

### 3. *Informationstheoretisch basierte didaktische Reduktion*

Beim Erlernen unserer Sprache lässt sich leicht errechnen, dass Kinder täglich etwa 3 neue Wörter lernen, was einem Wortschatz von rund 10.000 Wörtern in 10 Jahren entspricht. Dass wir diese rund 50.000 Byte durch Wiederholungen immer wieder in unserem Gedächtnis auffrischen müssen, liegt auf der Hand. Dabei herrscht zwischen dem Lernen und dem Vergessen ein steady state<sup>7</sup>, dass im Prinzip dem Postulat von Donald Hebb<sup>8</sup> (1904 – 1985) folgt, bei dem neuronale assoziative Verbindungen durch den jeweiligen Lernvorgang verbessert und stabilisiert werden. Dabei erreichen wir den Sprachzuwachs interessanterweise mit Hilfe des Zipfschen Gesetzes, das uns damit zwangsläufig zu dem führt, was

Abbildung 1: *Das Zipfsche Gesetz offenbart die Sprachökonomie, die uns unbewusst dazu führt, mit einem Minimum an Text, ein Maximum an Information zu übertragen. Die Unterscheidung zwischen Englisch ( $1/x$ ) und Deutsch ( $1/x^{0,8}$ ) ergibt sich aus den Charakteristika der jeweiligen Sprache, ist aber auch eine Frage des persönlichen Sprachgebrauchs. So können einzelne deutsche Texte durchaus auch die Potenz  $0,7$  ( $1/x^{0,7}$ ) erreichen.*



wir heute als informationstheoretisch basierte Didaktische Reduktion (IBDR) bezeichnen müssen. Worte, die wir täglich Hunderte mal lesen, schreiben oder sprechen, solche Worte lernen wir rascher, während wir die anderen um so langsamer lernen, je weniger sie gebraucht werden. Ihr Auftreten erscheint uns aber entsprechend der Seltenheit um so sensationeller, also informativer.

Zipf, G.K. hat in seinem Buch "Human Behavior and the Principle of Least Effort"<sup>9</sup> zwei Kräfte in unserer Sprache entdeckt. Die eine können wir auf Grund

- 7 Der Begriff Fließgleichgewicht wurde von dem Systemtheoretiker L. Bertalanffy für lebende Systeme eingeführt, die sich in einem stetigen Gleichgewicht zwischen Entropie und Negentropie (Information) befinden. Vgl. Bertalanffy, L. von, Biophysik des Fließgleichgewichts: Einführung in die Physik offener Systeme und ihre Anwendung in der Biologie. Hrsg.: Hermann Ebert. Braunschweig: Vieweg, 1953.
- 8 Hebb, D.O., The Organization of Behavior: A neuropsychological theory. New York: Wiley, 1949.

der Informationstheorie als Kraft in Richtung erhöhter Information und die andere in Richtung erhöhter Redundanz verstehen. Beide sind durch ein Potenzgesetz Verbunden, bei dem die Potenz im Englischen etwa 1 und im Deutschen eher 0,8 beträgt. Das bedeutet, dass die Kurve im Englischen etwa symmetrisch ist, während im Deutschen die Kraft zur Diversifikation ausgeprägter zu sein scheint. Vereinfacht gesagt ist im Englischen auf der einen Seite jedes 8te Wort THE und auf der anderen Seite, jedes 8te Wort, eines das im weiteren Text nicht mehr vorkommt.

Während wir im Deutschen durch die erhöhte Tendenz Worte zusammenzuschreiben mehr verschiedene Worte in einem Text haben, die nur ein einziges mal vorkommen, hat das Englische eine erhöhte Tendenz Worte getrennt zu schreiben und auch entsprechend oft benutzte Worte wie DER, DIE oder DAS sind im Englischen durch THE zusammengefasst. Wir folgen hier einem Sprachgefühl, dem bei genauer Betrachtung ein wichtiges informationstheoretisches Element der Optimierung des Verhältnisses von Information und Redundanz zugrunde liegt.

Interessanterweise zeigte die Rechtschreibreform an der zweiten Jahrtausendgrenze unserer Zeitrechnung so betrachtet eine deutliche Tendenz hin zur anglo-amerikanischen Sprachökonomie.

Wir können gemäß des Zipfschen Gesetzes, als Relation zwischen Information und Redundanz in der deutschen Sprache, auch schreiben:

$$\text{Redundanz} = \frac{\text{Konstante}}{\text{Information}^{0,8}}$$

Die Konstante kann darin als Schwellenwert verstanden werden, unter dem wir alles Rauschen ignorieren. Sie ist um so höher je umfangreicher der Text ist den wir betrachten. Das bedeutet, dass die Konstante kein absoluter Wert ist, sondern davon abhängt, wie viel Nachrichten auf uns einströmen. Dieser Datenstrom wird von uns beliebig kanalisiert, in dem wir beispielsweise nur mit einer bestimmten Geschwindigkeit lesen und bestimmte Sinne weitgehend schließen, wenn wir uns nur auf das Sehen, auf das Hören oder das Riechen konzentrieren wollen. Daraus folgt auch, dass es eine wirkliche Informationsflut nicht gibt. Im Gegenteil, der Mensch von heute ist mehr denn je auf der Suche nach Information, weil er im ständigen Wettbewerb mit seinen Mitstreitern und Gegnern steht.

Die sogenannte Informationsflut wird nur deshalb so empfunden, weil der Mensch sich seinen Wettbewerbern gegenüber gezwungen sieht, an die Grenzen

9 Zipf, G.K., *Human Behavior and the Principle of Least Effort; an introduction to human ecology*. Cambridge, Mass.: Addison-Wesley Press, 1949.

seiner Leistungsfähigkeit zu gehen. Wissenschaftler suchen heute mit Retrievalsystemen gezielt nach Informationen, überfliegen Texte in der Suche nach sensationellem und überprüfen beim sog. Zapping alle Kanäle von Fernsehen und Rundfunk wo etwas für sie wichtiges sein könnte – frei nach dem Motto F. Baccos, Wissen ist Macht. In der Wissenschaftsgesellschaft haben wir einen internationalen Konkurrenzkampf, in dem immer nur der Erste ein Patent oder Urheberrecht erwerben kann. Die Nahrungssuche in der Agrargesellschaft und die Rohstoffsuche in der Industriegesellschaft, weicht in der Wissenschaftsgesellschaft weitgehend der Suche nach Information und neuestem Wissen. Diesen immensen Druck von außen, oder genauer gesagt, den Informationssog, den wir permanent aufbauen um wettbewerbsfähig zu sein, empfinden wir als Informationsflut. Er ergibt sich aber eigentlich aus unserem Mangel an Wissen, den schon J. Naisbitt<sup>10</sup> in seinem Buch „Megatrends“ (1982) mit den bekannten Worten: „We are drowning in information and starved for knowledge“ beklagte.

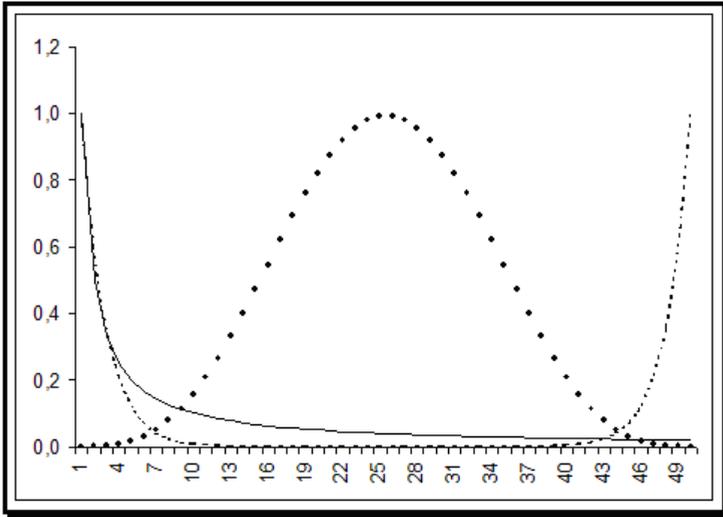
Soweit sich die Informationssuche auf sprachlicher Ebene abspielt, filtern wir entsprechend Zipfs „force of unification“ und „force of diversification“ nur die informationsreichsten, d.h. die außergewöhnlichen Worte, die uns auffallen, und bei den redundanzreichsten deren Häufigste heraus. Das sind sozusagen die Worte, die wir bevorzugt erlernen. Die gestrichelte Linie in Abb. 2 macht das zunächst dadurch deutlich, dass wir sozusagen mit abnehmender Worthäufigkeit auch Worte, die wir seltener gebrauchen leichter vergessen. Unsere Aufmerksamkeit steigt aber wieder rasch an, wenn wir Worte sehen oder hören, die so selten vorkommen, dass sie uns als Sensation erscheinen.

Die gepunktete Kurven in Abb. 2 macht deutlich, dass wir bei einem zunächst unbekanntem Text über das Luhnsche Prinzip herausfinden können wo die thematischen Schwerpunkte liegen, in dem wir ein Optimum zwischen Redundanz und Information in den extrahierten Worten suchen. Die Erkenntnis von H.P. Luhn (1958)<sup>11</sup> war, dass man aus Publikationen die Sätze herausfiltern kann, die vorwiegend Worte enthalten, die in einem gegebenen Text weder zu häufig (Stopwords wie der, die, das, ist, mit, und, er, es ...) noch zu selten (für den Text nicht typisch) auftraten. Die Exzerption solcher Sätze gibt stark abstrahiert das wahrscheinlichkeitstheoretische Charakteristische des Textes wieder. In diesem Fall filtern wir sozusagen invers, die Worte oder Wortverbindungen heraus, die weit informativer sind als die Stopwords, aber auch weit redundanter als die Wor-

10 Naisbitt, J., Megatrends. Ten New Directions Transforming Our Lives. New York: Warner Books, 1982.

11 Luhn, H.P., The automatic creation of literature abstracts, IBM Journal of Research and Development. 2(1958), S. 159 – 165.

*Abbildung 2: Während die durchgezogene Linie die Worthäufigkeitsverteilung nach Zipf darstellt, macht die punktierte ooooo Linie deutlich, welche Worte nach H. P. Luhn wahrscheinlich wichtig sind, um die thematische Charakteristik eines Textes zu erkennen, und die gestrichelte ---- Linie zeigt die Wahrscheinlichkeit, welche Worte beim Erlernen einer Sprache besonders wichtig sind.*



te, die nur ein mal im untersuchten Text erscheinen. Die Auswahl der Worte hängt somit einerseits von der Textlänge ab und andererseits von der Zipschen Verteilung von Redundanz und Information.

Beim Lesen bzw. Screening eines Textes gehen wir dem Zipschen Prinzip entsprechende so vor, dass wir die Zeilen überfliegen, und herausfiltern welche Worte uns wiederholt auffallen. Wobei wir aus der langjährigen Erfahrung heraus eine sehr klare Vorstellung davon haben, was Stopwords, sinntragende und extrem seltene Worte sind. Insofern tun wir das Selbe, was man beim Ranking in einer Datenbank tut, wenn man die Worthäufigkeit in einem Dokument, mit der in der gesamten Datenbank vergleicht.

Wenn die Suche eines bestimmten Wortes in einer Datenbank mit 1.000 Dokumenten beispielsweise zu 800 Treffern führt, und das Wort insgesamt 2.500 mal vorkommt, so ist dieses Wort für ein bestimmtes Dokument keinesfalls besonders charakteristisch, wenn es dort drei mal erscheint. Wenn ein anderes Wort dagegen in dem Dokument 5 mal erscheint, aber nur zu 50 Treffern führt, so

kann man durchaus davon ausgehen, dass dieses Wort ein Charakteristikum des Dokuments ist.

Dieses Wissen versetzt uns in die Lage, als Empfänger nicht nur auf Informationen zu warten, sondern aufgrund unseres Wissens dieses weitgehend vorauszusagen. Insofern ist die Erkenntnis keinesfalls erstaunlich, dass wir weitgehend nur das hören, sehen oder verstehen, was wir erwarten. Wenn wir zu einer Gruppe sich unterhaltender treten, müssen wir uns im allgemeinen kurz einhören, um bei jedem gesprochenen Satz schon zu ahnen, welche Wendung dieser Satz als nächstes nehmen wird. Schon allein die Tatsache, dass wir wissen, ob die Unterhaltung in deutsch oder englisch erfolgt, schränkt die zu erwartenden Worte erheblich ein.

Bei Erwachsenen beträgt die durchschnittliche Lesegeschwindigkeit 250 Worte pro Minute. Das entspricht, ähnlich wie beim sehen und hören etwa 170 Bit/sec. Eine getragene Ansprache, bei der der Redner verstanden werden will, hat aber kaum mehr als 100 Worte pro Minute und damit rund 80 Bit/sec. Berichte, nach denen es Menschen gibt, die die Lesegeschwindigkeit auf bis zu 10.000 Worte pro Minute (>6.000 Bit/s) zu erhöhen vermögen sind zwar denkbar, haben aber mit einem inhaltlichen Verstehen und verarbeiten dieser Textinhalte wenig zu tun. Das gehört eher in den Bereich des Screenings.

Im Gegenteil, es ist weitaus realistischer davon auszugehen, dass ein durchschnittlicher Wissenschaftler etwa 5.000 – 10.000 Worte pro Tag liest, und 100.000 überfliegt. Das führt zu 10.000 Publikationen, die er oder sie jährlich auf Brauchbarkeit überprüft, um etwa 100 Aufsätzen pro Jahr und 10 Bücher wirklich zu lesen. Dafür opfert er etwa 20% seiner Arbeitszeit. Bei Nichtwissenschaftlern dürften die überflogenen und gelesenen Texte durchaus in ähnlicher Größenordnung liegen, aber weniger wissenschaftlich sein. Damit ist die Vielfalt der benutzten Worte im Sprachgebrauch eher geringer.

Während die meisten Menschen einen Wortschatz in der Größenordnung von 10.000 pflegen, differenzieren Akademiker eher ein Mehrfaches davon. Das bedeutet, dass wenig wissenschaftlich interessierte Menschen sowohl beim Lesen als auch beim Sprechen einige wenige Worte Hunderte mal täglich verwenden, während Personen mit hoch differenziertem Sprachgebrauch durch die Wiederholung verschiedener spezieller Termini stärker zu präzisieren versuchen. Dabei wiederholen sie Fachbegriffe immer wieder in Abständen von Tagen, Wochen oder Monaten.

Die informationstheoretisch basierte Didaktische Reduktion kann man auch als Grundsatz des biologischen Lernens im Gegensatz zur herkömmlichen Didaktik verstehen, weil sie bei den Lernvorgängen aller Lebewesen beobachtbar ist, während die herkömmliche Didaktik, so wie wir sie in Schulen und anderen

Lehreinrichtungen kennen, diesen informationstheoretisch basierten Filtermechanismus der Didaktischen Reduktion im Prinzip vorweg nimmt, in dem die Schüler und Schülerinnen entsprechend der vorgegebenen Lehrpläne in ihren Schulbüchern weitgehend nur die Information erhalten, die sie sich erarbeiten sollen.

Die IBDR ist auch von fundamentaler Bedeutung für unser gesamtes Weltbild, weil sie uns darüber Auskunft gibt, was wir alles übersehen bzw. in dieser Welt ignorieren können. In erster Näherung gilt, je kleiner die Dinge in unserer Betrachtung erscheinen, desto eher achten wir auf die Redundanz. So erscheinen uns Mäuse und Ratten den Elefanten oder Blauwalen gegenüber eher erst in größerer Zahl bemerkenswert. Als Individuen sind sie aber grundsätzlich gleichgestellt. Sie kämpfen also ebenso um das Überleben ihres jeweiligen Genbestandes. Die Maus als Individuum wird erst bei genauem hinsehen und bei der Entdeckung individueller Eigenschaften informativ. Umgekehrt tritt auch die Individualität von Elefanten in großen Rudeln in den Hintergrund. Dieses geradezu radikale Filter der IBDR ist notwendig, weil wir nicht in der Lage sind, beispielsweise Tausend Mäuse als Einzelindividuen in ihrer Gesamtheit zu betrachten. Noch deutlicher wird dieses Phänomen, wenn wir die Individualität und Schönheit von Schneekristallen betrachten, die wir im allgemeinen nur als weiße Fläche sehen.

Wir suchen gemäß der IBDR in großen Gesamtheiten gewisse Gemeinsamkeiten, d.h. eine bestimmte Information und ihre Redundanz. Nach diesem Prinzip systematisieren wir unsere Welt und machen sie überschaubar.

In gewisser Hinsicht hat Jose Ortega y Gasset<sup>12</sup> in seinen „Betrachtungen über die Liebe“ diese als „ein Phänomen der Aufmerksamkeit“ angesehen. Unser „Bewusstsein verengt sich und enthält nur noch einen Gegenstand“, schreibt er, dem wir uns nach Platon durch „göttliche Besessenheit“ zuwenden. Sehr profan und wissenschaftlich betrachtet entledigen wir diesen speziellen Gegenstand unserer Liebe von jeder Redundanz und treiben seine Einmaligkeit sozusagen ins Unendliche.

Der Wert von Redundanz, also einer Information die wir bereits besitzen, und die bei ihrem Wiedereintreffen nur bereits bekanntes enthält, ist, so lässt sich leicht vermuten, relativ gering, denn diese Redundanz bringt uns ja nichts wirklich neues. Trotzdem ist dieser Wert nicht Null, weil die Redundanz uns eine erhöhte Sicherheit dafür gibt, dass die dazugehörige Information, die wir vorher erhielten, an Verlässlichkeit gewinnt.

Wenn wir beispielsweise hören, dass es draußen kalt ist, so nehmen wir dies zunächst zur Kenntnis, ohne zu wissen, ob die Aussage stimmt. Wenn wir diese

12 Ortega y Gasset, J., Betrachtungen über die Liebe. Frankfurt am Main: Suhrkamp, 1991, 1. Aufl.

Botschaft, dagegen ein zweites oder drittes mal empfangen, nehmen wir sie berechtigterweise als wahrscheinlich richtig an. Dabei registrieren wir durchaus, ob die Wiederholung vom selben Sender ausgegangen ist, oder uns unabhängig voneinander erreicht.

Sobald uns Nachrichten unabhängig voneinander erreichen, also beispielweise von verschiedenen Personen überbracht werden, die von draußen hereinkommen, oder auch von verschiedenen Thermometern die sich draußen befinden, so wächst die Zuverlässigkeit der Nachricht natürlich stärker, als bei einem wiederholten Empfang von der selben Quelle. Bezogen auf das Wissen in Bibliotheken bedeutet dies, eine erhöhte Sicherheit, wenn verschiedene Autoren auf unterschiedlichen Wegen zum selben Wissen gelangen.

#### 4. *Bewertung von Information und Wissen*

Wir müssen bei der Bewertung von Information zunächst zwei verschiedene Bereiche, hinsichtlich des Nutzens den sie uns bringt, unterscheiden.

1. Welchen Nutzen bringt uns die Information selbst.
2. Welchen Nutzen bringt uns die Zuverlässigkeit dieser Information

Für den ersten Bereich gibt es zahlreiche Beispiele, wie die Informationen über Verdienstmöglichkeiten, Sonderangebote, freie Stellen, über Energieeinsparungsmöglichkeiten oder technische Hilfsmittel zur Arbeitserleichterung etc.

Der zweite Bereich ist weitgehend von unserem Wissen abhängig, da Wissen als eine besondere Form der Redundanz, als a priori Redundanz, anzusehen ist. Herkömmliche Redundanz ist eine a posteriori Redundanz, weil wir sie erst aus der vorhergegangenen Information erkennen. Bei der a priori Redundanz erwarten wir auf Grund unseres Wissens eine bestimmte Information. Tritt diese ein, so halten wir sie für wahrscheinlich richtig. Dies ist auch der Grund, warum Wissen sich nicht nur, wie oft angenommen selbst organisiert, es reproduziert sich selbst, so dass verschiedene Denkvorgänge bzw. verschiedene Wissenschaftler, überall dort wo wir Wissen erringen können, zu den selben Ergebnissen gelangen. Anderenfalls gäbe es keine Wissenschaft. Darin liegt auch der Grund, dass sich Wissen weitgehend logisch aus seinen Urgründen heraus selbst ergibt.

Bezogen auf das Beispiel Sonderangebote haben wir abzuschätzen, ob eine Ware für einen besonders niedrigen Preis ein Schnäppchen, Tinnef oder aber Betrug ist. Wir rekonstruieren, soweit das möglich ist, aus Erfahrung und Logik ihren wahrscheinlichen Warenwert und den sich daraus ergebenden Marktpreis.

Die Bewertung von Information hat nichts mit seiner Messbarkeit zu tun. Ebenso wenig, wie ein Auto, nach seiner Länge oder seinem Gewicht bewertet

werden kann, wenn es um Komfort oder um die Geschwindigkeit geht. Der finanzielle Wert ist beispielsweise von seiner Einsetzbarkeit abhängig, und davon, wie viel Geld man damit einsparen oder auch verdienen kann. In gleicher Weise, sagt die Menge an Wissen in einer Bibliothek zunächst nichts über ihren Marktwert oder ihren ideellen Wert aus.

Während im Jahre 2000 noch etwa 1,5 Exabyte, bzw. 250 Megabyte pro Erdbewohner anfielen, waren es 2002, nach Schätzungen einer Berkeley-Studie schon 5 Exabyte oder 5 Mio. Terabyte ( $4 \times 10^{19}$  Bit).<sup>13</sup>

Bei einer Weltbevölkerung von 6,3 Milliarden Menschen entfallen auf jeden Erdbewohner etwa 800 Megabyte an neuen Daten. Das entspricht einer Verdopplungsrate von 1,1 Jahren. Dabei darf allerdings nicht vergessen werden, wie viel Redundanz, Rauschen und in semiotischer Hinsicht auch wie viel Unsinn mit gespeichert wird. Außerdem sind große Teile dessen Bild und Tonmaterialien, die sehr Speicherintensiv sind.

Dreiundneunzig Prozent der neuen Daten sind digital gespeichert worden, davon über die Hälfte auf Festplatten, 7 Prozent auf Filmen und nur ein Bruchteil auf Papier. Nordamerikaner nutzen etwa 12.000 Blatt Papier / Person und Jahr, während Europäer nur auf 5.000 kommen. Die USA produzieren insgesamt 40 Prozent der gespeicherten Informationen. Verglichen mit den Informationsmengen der Library of Congress sind das so große Zahlen, dass die LC nur noch im Promille-Bereich erscheint. Es wird geschätzt, dass heute nur noch 0,03 % der Dokumente in gedruckter Form vorliegen. Wenn man dabei bedenkt, dass viele Fachleute noch eine Archivierung auf Papier für denkbar halten, so zeigt sich rasch die Abwegigkeit solcher Überlegungen. Andererseits ist aber der Anteil an Wissen in der LC sicher höher, als in dieser Gesamtschau jährlich hinzukommender Daten, von denen allein 80 Mrd. Fotografien pro Jahr zu Buche schlagen. Diese Bilder enthalten informationstheoretisch fast kein Wissen, auch wenn ihr Informationsgehalt meist um ein vielfaches größer ist, als er beispielsweise in wissenschaftlichen Texten mit ihren Begründungen sein kann.

13 Lyman, P. / Varian, H.R. / Dunn, J. / Strygin, A. / Swearingen, K., How much information. 2003. <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm>

## 5. *Schluss*

Wissen ist die höchste Form der Informationskompression die uns zur Verfügung steht. Nur wenige Promille dessen was wir an publizierter Information besitzen, betrifft das Wissen der Menschheit. Vieles gehört noch zum tacit knowledge und zu dem was wir noch nicht publiziert haben. Zieht man die Menge an gespeichertem Wissen in der Biologie zum Vergleich heran, so wird deutlich, dass die exponentielle Skalierung von Wissen und Information es durchaus denkbar erscheinen lässt, dass wir das Menschliche Wissen in Bereichen von weniger als 1010 Bit zu komprimieren vermögen. Dabei ist zu berücksichtigen, dass Wissen durch fünf Eigenschaften charakterisiert ist:

1. durch die Präzision
2. durch die Verlässlichkeit
3. durch die Reichweite in der Zeit
4. durch die thematische Reichweite
5. und durch die Kompression.

Je nach Wissensumfang, Präzision und Reichweite kann die Kompression wirksam werden. Damit hat Wissen sozusagen 4 Dimensionen, in denen bestimmt wird, wie weit eine erwartete Information mit der real empfangenen übereinstimmt.

Die erste Dimension betrifft die Klassierung, aus der sich ergibt, ob wir eine eintreffende Information noch als richtig tolerieren können oder nicht. Werten wir die Aussage, morgen geht die Sonne um 6 Uhr auf als richtig, wenn sie um drei Minuten vor sechs aufgeht?

Die zweite Dimension betrifft die Zufälligkeit, die wir berücksichtigen müssen. Wie bewerten wir Aussagen, die in 85 von 100 Fällen als richtig eingestuft werden?

Die dritte Dimension betrifft die Frage, ob unsere Aussage für eintreffende Informationen in den nächsten Sekunden, Tagen oder Jahrzehnten gilt. Dabei muss berücksichtigt werden, dass wir eine Aussage oder auch ein Messergebnis in exakt zwei Minuten erwarten können, das aber die Vorhersage eines Ereignisses betrifft, das vor Millionen Jahren geschah. Ebenso können wir z.B. eine Wahlvorhersage für den jetzigen Moment, in dem eine Wahl abgeschlossen wurde, tätigen, von der wir aber zunächst nicht wissen, wann das endgültige Wahlergebnis veröffentlicht werden wird.

Die vierte Dimension betrifft die Frage, ob beispielsweise eine Vorhersage für Bevölkerungsentwicklungen, aufgrund eines Modells, nur für eine Region, ein Land, die Menschheit oder für alle Lebewesen gilt.

Bei all diesen Fragen müssen wir aber immer berücksichtigen, wie weit auf einem bestimmten Gebiet überhaupt Wissen erworben werden kann. Denn auch hier gilt die Gaußsche Erkenntnis: „Der Mangel an mathematischer Bildung gibt sich durch nichts so auffallend zu erkennen, wie durch maßlose Schärfe im Zahlenrechnen.“<sup>14</sup> Insofern gehört es essentiell zur Messung von Wissen, dessen Begrenztheit mit einzuschätzen. Das betrifft sowohl die Bereiche, wie die Heisenbergsche Unschärferelation, in denen wir sozusagen objektiv die Grenze unseres Wissens angeben können, als auch unsere persönlichen bzw. subjektiven Wissensgrenzen.

Die Kompression ist selbst keine eigene Dimension des Wissens, sondern ein Ergebnis der Skalierung von Information, Redundanz und Wissen. Diese exponentielle Skalierung führt dazu, dass wir mit wachsendem Wissen immer mehr Informationen in ein in sich logisches Gebäude, unser Weltbild, einfügen können, so dass die Menschheit, ohne einen gravierenden Mehrbedarf an Erinnerungsvermögen, wir könnten auch sagen ohne signifikant erhöhten Speicherbedarf, immer mehr Wissen in unser Bewusstsein aufnehmen können.

Darin liegt auch der Grund, dass wir unser Wissen immer wieder neu publizieren und zusammenfügen, so dass wir zwar in unseren Bibliotheken immer wieder nach alten Informationen suchen, die wir noch nicht in unser Weltbild sinnvoll einbeziehen konnten, während wir das Wissen, insbesondere in der Digitalen Bibliothek, immer wieder neu reorganisieren und damit auch komprimieren.

14 Motto in Küster, F.W. / Thiel, F., *Logarithmische Rechentafeln für Chemiker, Pharmazeuten, Mediziner und Physiker*, neu bearbeitet von K. Fischbeck. Walter de Gruyter Co. Berlin 1947



*The Part of Knowledge in Libraries.**(Der Anteil an Wissen in Bibliotheken)**Walther Umstätter*

In 1963 the amount of information was estimated at first time with 10<sup>13</sup> Bit. This was a very rough assessment including two errors. On one hand it regarded only textual parts of the books and neglected all pictures, tables etc., and on the other hand it ignored the great amount of redundancy in such a library. The question about the amount of knowledge in such a library, in the published world, in human brains or in the genomes of animals and plants is of high importance, if we want to understand the difference between information and knowledge management. In this context it is remarkable, that knowledge is in some areas a very effective compression of information, because it is not only self-organizing, it is also self-reproductive. That means, that we can have a very small basis of knowledge for millions of explanations. That is the cause, that measurement of knowledge is not based on a linear scale but on an exponential scale.

---

Gesellschaft für  
Wissenschaftsforschung



Klaus Fuchs-Kittowski,  
Walther Umstätter  
Roland Wagner-Döbler (Hrsg.)

Wissensmanagement  
in der Wissenschaft

Wissenschaftsforschung  
Jahrbuch 2004

Mit Beiträgen von:

*Wladimir Bodrow • Klaus Fuchs-  
Kittowski Jay Hauben • Matthias Kölbel •  
Peter Mambrey • Erhard Nullmeier •  
Walther Umstätter • Rose Vogel • Sven*

Wissenschaftsforschung **2004**  
Jahrbuch

---

**Wissensmanagement in der Wissenschaft:**

Wissenschaftsforschung Jahrbuch 2004 / Klaus  
Fuchs-Kittowski; Walther Umstätter; Roland  
Wagner-Döbler (Hrsg.). Mit Beiträgen von  
Wladimir Bodrow ... - Berlin: Gesellschaft für  
Wissenschaftsforschung 2005.

Bibliographische Informationen Der Deutschen  
Bibliothek

Die Deutsche Bibliothek verzeichnet diese  
Publikation in der Deutschen  
Nationalbibliographie; detaillierte  
bibliographische Daten sind im Internet über  
<http://dnb.ddb.de> abrufbar.

Das Werk ist in allen seinen Teilen urheberrechtlich  
geschützt.

Jede kommerzielle Verwertung ohne schriftliche  
Genehmigung des Verlages ist unzulässig. Dies gilt  
insbesondere für Vervielfältigungen, Übersetzungen,  
Mikroverfilmungen und die Einspeicherung und  
Verarbeitung in Systeme(n) der elektronischen  
Datenverarbeitung.

© Gesellschaft für Wissenschaftsforschung,  
1. Auflage 2005  
Alle Rechte vorbehalten.

Verlag:  
Gesellschaft für Wissenschaftsforschung  
c/o Prof. Dr. Walther Umstätter, Institut für  
Bibliothekswissenschaft der Humboldt-Universität zu  
Berlin, Dorotheenstr. 26, D-10099 Berlin

Druck: BOOKS on DEMAND GmbH,  
Gutenbergring, D-22848 Norderstedt

ISBN 3-934682-39-1  
Preis: 15,80 €