It from bit: the science of information

Richard Harvey
IT Livery Company Professor of Information Technology, Gresham College
Professor of Computer Science, School of Computing Sciences, University of East Anglia

#richardwharvey
information | ɪnfəˈmeɪʃ(ə)n |

noun [mass noun]

1 facts provided or learned about something or someone: a vital piece of information.

   [count noun] Law a charge lodged with a magistrates' court: the tenant may lay an information against his landlord.

2 what is conveyed or represented by a particular arrangement or sequence of things: genetically transmitted information.

   Computing data as processed, stored, or transmitted by a computer

   (in information theory) a mathematical quantity expressing the probability of occurrence of a particular sequence of symbols, impulses, etc., as against that of alternative sequences.

ORIGIN late Middle English (also in the sense ‘formation of the mind, teaching’), via Old French from Latin informatio(n-), from the verb informare
information | ɪnfəˈmeɪʃ(ə)n |

noun [mass noun]

1 facts provided or learned about something or someone: a vital piece of information.

   [count noun] Law a charge lodged with a magistrates' court: the tenant may lay an information against his landlord.

2 what is conveyed or represented by a particular arrangement or sequence of things: genetically transmitted information.

   Computing data as processed, stored, or transmitted by a computer

   (in information theory) a mathematical quantity expressing the probability of occurrence of a particular sequence of symbols, impulses, etc., as against that of alternative sequences.

ORIGIN late Middle English (also in the sense ‘formation of the mind, teaching’), via Old French from Latin informatio(n-), from the verb informare
Shannon information
Artefacts from the MIT Museum
https://mitmuseum.mit.edu
A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist and Hartley on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonically increasing function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance

would produce “messages” consisting of a number of functions of three variables; (f) Various combinations also occur, for example in television with an associated audio channel.

2. A transmitter which operates on the message in some way to produce a signal suitable for transmission over the channel. In telephony this operation consists merely of changing sound pressure into a proportional electrical current. In telegraphy we have an encoding operation which produces a sequence of dots, dashes and spaces on the channel corresponding to the message. In a multiplex PCM system the different speech functions must be sampled, compressed, quantized and encoded, and finally interleaved
In economics, Gresham's law is a monetary principle stating that "bad money drives out good". For example, if there are two forms of commodity money in circulation, which are accepted by law as having similar face value, the more valuable commodity will gradually disappear from circulation.

The law was named in 1860 by Henry Dunning Macleod, after Sir Thomas Gresham (1519–1579), who was an English financier during the Tudor dynasty. However, there are numerous predecessors. The law had been state...
How to code this text?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>0 1 0 0 0 0 0 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0 1 0 0 0 0 1 0</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0 1 0 0 0 1 1 1</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0 1 0 0 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0 1 0 0 1 0 1 1</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0 1 0 0 1 1 0 0</td>
<td></td>
</tr>
</tbody>
</table>

500 characters $\times$ 8 bits per character = 4000 bits
But some letters are more common than others
If Youth, throughout all history, had had a champion to stand up for it; to show a doubting world that a child can think; and, possibly, do it practically; you wouldn't constantly run across folks today who claim that “a child don't know anything.” A child's brain starts functioning at birth; and has, amongst its many infant convolutions, thousands of dormant atoms, into which God has put a mystic possibility for noticing an adult's act, and figuring out its purport. …
So what about a different code?

Maybe we could use fewer bits for the most common letters?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th></th>
<th>L_A = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>L_B = 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>L_C = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>L_D = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>L_E = 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>L_F = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average length of this code is

\[ p_A L_A + p_B L_B + \ldots + p_Z L_Z \]

= 4.2 bits

Roughly half the size of ASCII
Information $I = - \log_2 p$
So what’s the best we could do?

Shannon Entropy

\[ H = - \sum_{n=1}^{N} p_n \log_2 p_n = \text{Average}( -\log_2 p) \]

For our probabilities:

\[ H = 4.176 \text{ bits per character.} \]
Improbability = surprise = information = bits
Information (and surprise) is conditional

Given that I know X, how much extra information is needed to describe Y?
Information is conditional
Surprise is conditional
Surprise is conditional
Magic compression

A *magic* or universal compressor would be able to compress every file...

If such a thing existed then it would be able to compress files that themselves were compressed files

So the universal compressor does not exist!
Lossless compression

- **Text**
  - PAQ8PXD_v47

- **Video**
  - FFV1 Lossless -c:v ffv1 -level 3 -pass 1 -coder 1 -context 1

- **Images**
  - Emma 0.1.23

- **Audio**
  - SAC 0.0.6a4 --insane

[Lossless compression chart]
Video data

1080 lines × 1920 columns = 2M pixels
2M pixels × 3 colours × 8 bits = 49.8 Mbits per frame
49.8 Mbits per frame × 25 frames per second = 1.2 Gbits per second

But DVB channels range from 5 Mbits per second to 50 Mbits per second.

Compression ratios of 250 to 25 needed.
Sound is not what meets the ear

For example: https://www.youtube.com/watch?v=BzNzgsAE4F0
I wonder if this is encoded for humans?

By Francis Barraud
1415926535...

$I = -\log_2 (1/10) = 3.3$ bits per symbol

10 symbols = 33 bits
14159265358979323846264338327
95028841971693993751058209749
44592307816406286208998628034
825342117067...

100 symbols = 330 bits
1000 symbols = 3300 bits
...and so on...

1 million symbols = 3.3 Mbits
...but...
1415926535...

are the first 10 digits of...

3.1415926535... = π
KOLMOGOROV INFORMATION

The length of the shortest program that can compute the object
KOLMOGOROV INFORMATION

The length of the shortest program that can compute the object
160 character program for \( \pi \):

```c
int a=10000, b, c=2800, d, e, f[2801], g;
main(){for(;b-c;)f[b++]=a/5;
    for(;d=0, g=c*2; c--=14, printf("%.4d", e+d/a), e=d%a) for(b=c; d+=f[b]*a,
        f[b]=d%--g, d/=g--, --b; d*=b);}
```

\( \pi \) can be computed with a 160 character (1280 bit) program

so the information in \( \pi \) is < 1280 bits
KOLMOGOROV INFORMATION IS CONDITIONAL TOO

$K(x|y)$ is the size of the shortest program that computes $x$ given $y$ as input

A string is incompressible if $K(x) = \text{Length}(x)$
INFORMATION = SIZE OF MACHINE THAT CREATES IT
THING  ←  PATTERN

ALGORITHM  ←  CODE
OLD SCIENCE = WHAT IS THAT THING?

NEW SCIENCE = WHAT CREATES THAT THING?
"Generation of Leaf Shape Through Early Patterns of Growth and Tissue Polarity",
Erika E. Kuchen, Samantha Fox, Pierre Barbier de Reuille, Richard Kennaway, Sandra Bensmihen, Jerome Avondo, Grant M. Calder, Paul Southam, Sarah Robinson, Andrew Bangham, Enrico Coen, Science, 335, 2 March 2012
IT FROM BIT?
Unanswered questions

- digital physics
- philosophy
- information theory
- applications
- technology
Moving onwards:
speech (27th Nov)
vision (12th Feb)
learning (19th March)
text (16th April)
creativity (28th May)