The Economist

All too much

Monstrous amounts of data

Feb 25th 2010 | from the print edition

QUANTIFYING the amount of information that exists in the world is hard. What is clear is that there is an awful lot of it, and it is growing at a terrific rate (a compound annual 60%) that is speeding up all the time. The flood of data from sensors, computers, research labs, cameras, phones and the like surpassed the capacity of storage technologies in 2007. Experiments at the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, generate 40 terabytes every second—orders of magnitude more than can be stored or analysed. So scientists collect what they can and let the rest dissipate into the ether.

According to a 2008 study by International Data Corp (IDC), a market-research firm, around 1,200 exabytes of digital data will be generated this year. Other studies measure slightly different things. Hal Varian and the late Peter Lyman of the University of California in Berkeley, who pioneered the idea of counting the world's bits, came up with a far smaller amount, around 5 exabytes in 2002, because they counted only the stock of original content.

Data inflation 2

Unit	Size	What it means
Bit (b)	1 or 0	Short for "binary digit", after the binary code (1 or 0) computers use to store and process data
Byte (B)	8 bits	Enough information to create an English letter or number in computer code. It is the basic unit of computing
Kilobyte (KB)	1,000, or 2^{10}, bytes	From "thousand" in Greek. One page of typed text is 2KB
Megabyte (MB)	1,000KB; 2^{20} bytes	From "large" in Greek. The complete works of Shakespeare total 5MB. A typical pop song is about 4MB
Gigabyte (GB)	1,000MB; 2^{30} bytes	From "giant" in Greek. A two-hour film can be compressed into 1-2GB
Terabyte (TB)	1,000GB; 2^{40} bytes	From "monster" in Greek. All the catalogued books in America's Library of Congress total 15TB
Petabyte (PB)	1,000TB; 2^{50} bytes	All letters delivered by America's postal service this year will amount to around 5PB. Google processes around 1PB every hour
Exabyte (EB)	1,000PB; 2^{60} bytes	Equivalent to 10 billion copies of *The Economist*
Zettabyte (ZB)	1,000EB; 2^{70} bytes	The total amount of information in existence this year is forecast to be around 1.2ZB
Yottabyte (YB)	1,000ZB; 2^{80} bytes	Currently too big to imagine

Source: *The Economist*

The prefixes are set by an intergovernmental group, the International Bureau of Weights and Measures. Yotta and Zetta were added in 1991; terms for larger amounts have yet to be established.

What about the information that is actually consumed? Researchers at the University of California in San Diego (UCSD) examined the flow of data to American households. They found that in 2008 such households were bombarded with 3.6 zettabytes of information (or 34 gigabytes per person per day). The biggest data hogs were video games and television. In terms of bytes, written words are insignificant, amounting to less than 0.1% of the total. However, the amount of reading people do, previously in decline because of television, has almost tripled since 1980, thanks to all that text on the internet. In the past information consumption was largely passive, leaving aside the telephone. Today half of all bytes are received interactively, according to the UCSD. Future studies will extend beyond American households to quantify consumption globally and include business use as well.

March of the machines

Significantly, "information created by machines and used by other machines will probably grow faster than anything else," explains Roger Bohn of the UCSD, one of the authors of the study on American households. "This is primarily 'database to database' information—people are only tangentially involved in most of it."

retrievable and usable. But this is changing as content on the web is increasingly “tagged”, and facial-recognition and voice-recognition software can identify people and words in digital files.

“It is a very sad thing that nowadays there is so little useless information,” quipped Oscar Wilde in 1894. He did not know the half of it.

from the print edition | Special reports