



Information

quantity and growth

5 exabytes of new information in 2002

- Print, film, magnetic, and optical storage media produced about 5 exabytes of new information in 2002
 - How big is 5 exabytes?
 - If digitized with full formatting, the seventeen million books in the Library of Congress contain about 136 terabytes of information
 - five exabytes of information is equivalent in size to the information contained in 37 000 libraries of the size of the Library of Congress book collections.
- new stored information grew about 30% a year between 1999 and 2002.
- Instant messaging: 5 billion messages/day (750GB), or 274 Terabytes/year.
- Email: 400,000 terabytes of new information each year worldwide.

Information: measure

■ Gigabyte (GB) 1 000 000 000 bytes OR 10^9 bytes

- ★ 1 Gigabyte: a pickup truck filled with books.
- ★ 100 Gigabytes: A library floor of academic journals.

■ Terabyte (TB) 1 000 000 000 000 bytes OR 10^{12} bytes

- ★ 1 Terabyte: 50000 trees made into paper and printed.
- ★ 2 Terabytes: An academic research library.
- ★ 10 Terabytes: The print collections of the U.S. Library of Congress.

■ Petabyte (PB) 1 000 000 000 000 000 bytes OR 10^{15} bytes

- ★ 2 Petabytes: All U.S. academic research libraries.
- ★ 20 Petabytes: Production of hard-disk drives in 1995.
- ★ 200 Petabytes: All printed material.

■ Exabyte (EB) 1 000 000 000 000 000 000 bytes OR 10^{18} bytes

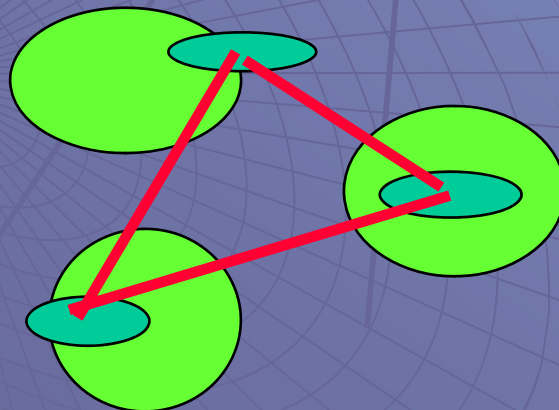
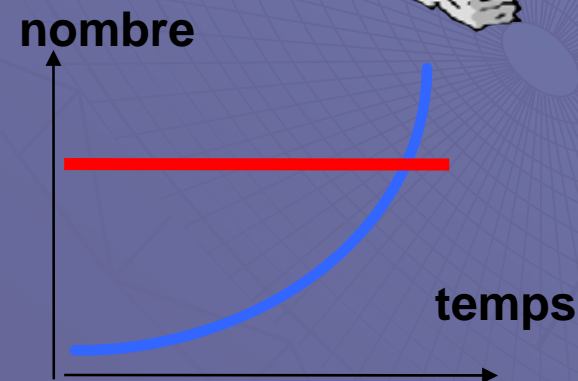
- ★ 2 Exabytes: Total volume of information generated in 1999.
- ★ 5 Exabytes: All words ever spoken by human beings.

It is still going on!!

- In spite of 2008 crisis,
 - the augment of numerical information accelerate (IDC)
 - 3 892 179 868 480 350 000 000
 - 3 trilliards 892 trillions 179 billiards 868 billions 480 milliards 350 millions of new digital information bits were created and transmit by Internet, phone networks, waves during 2008.

Human brain limitation

- 2 Kilobytes: A Typewritten page.
- 30 Megabytes: 30 cm of shelved books: the amount of memorized information/year by an expert
- 500 pages book 1 Mb
- Quando se fala de busca da informação: 1^{ero} reflexo e o Google



$$1+1+1 > 3$$

Evaluation of the visible web

- Google.com 16.400.000.000 pages (May 2008)
 - X 474 words (average) = 7.77×10^{12} Words
 - X 41 links (average) = 672 400 000 000 links
 - 2,7 billions searches each month

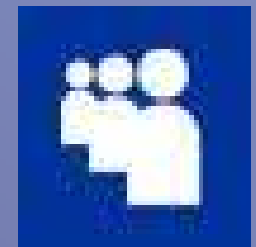


Wikipedia

- 1.74 billion words, 9.25 million articles, 250 languages
- English Wikipedia: over 1 billion words, 25 times the largest English-encyclopedia
- 282.885 contributors (151.937 to the English language edition)



- Myspace population over 125.000.000



- Shakespeare had 24 000 words to work with...

Invisible or Deep or Hidden Web

- Surface Web: indexed by common search engines
 - 170 terabytes of information on its surface
 - 17 times the size of the Library of Congress print collections.
- Invisible or deep or hidden Web
 - pages on the World Wide Web that are not part of the surface web
 - pages not linked to by other pages, such as Dynamic Web pages.
 - searchable databases that deliver Web pages generated just in response to a query and contain information stored in tables created by programs such as Access, Oracle or SQL databases.
 - sites that require registration or limit access to their pages, prohibiting search engines from browsing them and creating cached copies.
- 500 times larger than the content accessible via the surface Web



