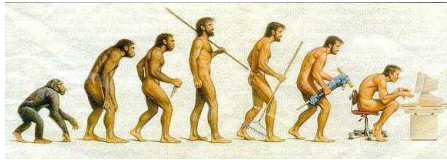


# I documenti digitali



## Obiettivi della lezione

- Introduzione ai documenti digitali
- I nuovi canali di comunicazione
- Il ciclo di vita dei documenti digitali
- L'industria dei contenuti digitali

## Il futuro dell'editoria

Last Updated: Thursday, 3 November 2005, 13:34 GMT  
[E-mail this to a friend](#) [Printable version](#)

### First Google Print books launched

**The first group of works scanned and put online by the Google Print project has been unveiled.**

Included in this opening swathe are many 19th Century works of American literature and history.

The works were chosen because they are uncopyrighted works that are unaffected by the legal action that briefly halted Google's digitisation project.

The legal challenge comes from the Authors Guild which claims it involves "massive copyright infringement".

#### Legal row

Books about the US Civil War, government papers and the writings of Henry James are among the works donated by project participants for the first group of online works.



The first works from Google Print's project have been unveiled

#### SEE ALSO:

- [Microsoft aims to trounce Google](#) 27 Oct 05 | Technology
- [Ads fuel soaring Google profits](#) 20 Oct 05 | Business
- [Google restarts online books plan](#) 01 Nov 05 | Technology
- [Google is sued by book publishers](#) 19 Oct 05 | Business
- [Microsoft joins book search plan](#) 26 Oct 05 | Technology
- [Yahoo backs digital library plan](#) 03 Oct 05 | Technology

#### RELATED INTERNET LINKS:

- [Authors Guild on Google lawsuit](#)
- [US Authors Guild](#)
- [Google](#)
- [Google Print](#)
- [Open Content Alliance](#)

The BBC is not responsible for the content of external internet sites

**TOP TECHNOLOGY STORIES NOW**



**One Hundred Years of Solitude**  
by Gabriel García Márquez  
Provided by [HarperCollins](#) through the [Google Print Publisher Program](#)

Page 2

[Back to Search results](#)



[HarperCollins Publishers](#)

Search within this book

the world was so recent

[More results from this book](#)

Book Pages

[About this Book](#)  
[Google](#)  
[Back Cover](#)

Buy this Book

[Amazon.com](#)  
[BarnesAndNoble.com](#)  
[BookSense.com](#)  
[Fisade](#)

[About Google Print](#)

2 Gabriel García Márquez

appeared from where they had been searched for most and went dragging along in turbulent confusion behind Melquiades' magical irons. "Things have a life of their own," the gypsy proclaimed with a harsh accent. "It's simply a matter of waking up their souls." José Arcadio Buendía, whose unbridled imagination always went beyond the genius of nature and even beyond miracles and magic, thought that it would be possible to make use of that useless invention to extract gold from the bowels of the earth. Melquiades, who was an honest man, warned him: "It won't work for that." But José Arcadio Buendía at that time did not believe in the honesty of gypsies, so he traded his mule and a pair of goats for the two magnetized ingots. Úrsula Iguarán, his wife, who relied on those animals to increase their poor domestic holdings, was unable to dissuade him. "Very soon we'll have gold enough and more to pave the floors of the house," her husband replied. For several months he worked hard to demonstrate the truth of his idea. He explored every inch of the region, even the riverbed, dragging the two iron ingots along and reciting Melquiades' incantation aloud. The only thing he succeeded in doing was to unearth a suit of fifteenth-century armor which had all of its pieces soldered together with rust and inside of which there was the hollow res-

Copyrighted Material

## Microsoft digitalizza la British Library

**Il gigante del software vuole portare in rete le opere della biblioteca britannica, Amazon lancia una nuova formula di vendita**



Una preziosa edizione del Vangelo conservata dalla British Library (Reuters)

**LONDRA** – All'indomani del debutto di Google-Print, la contestata internet library del motore di ricerca di Mountain View, si parla ancora di grandi novità nel settore dell'editoria in formato elettronico.

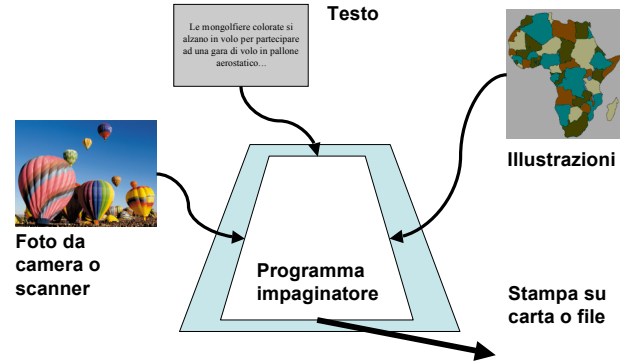
**LA BIBLIOTECA BRITANNICA** - Questa volta al centro dei riflettori c'è Microsoft, con l'annuncio che dal 2006 si impegnerà in un'impegnativa opera di digitalizzazione di circa 100 mila testi (per un totale di 25 milioni di pagine) conservati nei preziosi archivi della British Library. Una volta convertite in formato elettronico le opere saranno messe online a disposizione del pubblico di tutto il mondo, che potrà accedere alla nuova biblioteca attraverso il servizio di ricerca libri offerto dal colosso americano del software. L'iniziativa della società di Bill Gates rientra nell'ambito dell'importante progetto di digitalizzazione di libri promosso dalla Open Content Alliance - a cui aderiscono, tra gli altri, anche Yahoo, HP e Adobe - e prevede che molti altri libri entrino presto a far parte della nuova collezione consultabile via web.

**LIBRI A CONSUMO DA AMAZON** - Nel frattempo il primo book retailer del web lancia una nuova formula di vendita, proponendo ai suoi utenti una sorta di acquisto "a consumo". In pratica, il programma Amazon Pages permette ai lettori di comprare singole pagine o singoli capitoli dei libri in formato digitale presenti sui suoi cataloghi. Il prezzo dipenderà dal tipo di libro e dall'editore, ma in linea di massima sarà di pochi centesimi per pagina.

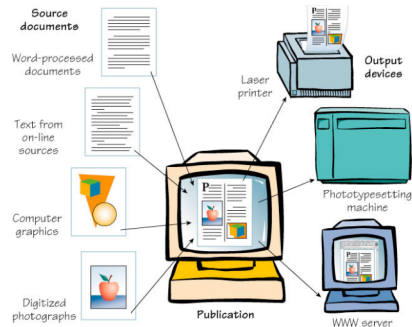
**Alessandra Carboni**

04 novembre 2005

## I documenti digitali sono oggetti "costruibili"



## I documenti digitali sono distribuibili su più canali



## Nuovi canali di comunicazione

- Siti web
- Portali es. [www.unibo.it](http://www.unibo.it)
- Blog es. [www.technorati.com/pop/blogs/](http://www.technorati.com/pop/blogs/)
- Forum es. [www.filmbug.com/site/topforums.php](http://www.filmbug.com/site/topforums.php)
- Wiki es. [en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)
- Really Simple Syndication (RSS)
- Motori di ricerca

## Documento [Devoto Oli]

- Scrittura atta a fornire un'autorevole prova o convalida
- Attestato rilasciato da pubblica autorità
- Qualsiasi oggetto utilizzabile ai fini di consultazione, ricerca, informazione
- Illustrazione, dimostrazione

## Document [Webster]

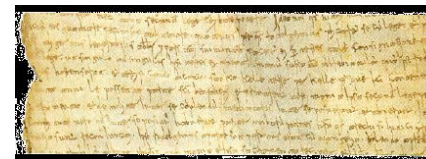
- Anything written, printed, etc., relied upon to record or prove something
- Anything serving as a proof
- Dal latino *documentum*: lezione, esempio, prova

## Documento

Documento:

**Contenitore** (o *supporto*) di dati (testo, numeri, figure, ecc.) strutturati per essere usati come informazioni

## Un documento



**Sao ko kelle terre, per kelle fini que ki contene,  
trenta anni le possette parte sancti Benedicti...**

Placito Capuano: *atto giudiziario custodito nel monastero di Montecassino, primo scritto ufficiale in lingua volgare (Capua, marzo 960)*

## Documento digitale

### Documento digitale:

Documento rappresentato in forma **binaria** e **memorizzato** in un sistema informatico

Ha un **nome**, un **contenuto**, una **struttura**, alcuni **comportamenti**, alcune **relazioni** con altri documenti

Il contenuto è di solito **codificato** (anche più volte, da codici sovrapposti)

## Codici sovrapposti

- I documenti digitali che contengono testo sono rappresentati internamente ad un sistema informatico:
  - da un **codice alfanumerico**, ad esempio ASCII o Unicode, che definisce la rappresentazione del testo del documento in forma di bit
  - da un **codice di markup**, ad esempio il .doc di Word, che definisce la *struttura* del documento in funzione dell'applicazione che lo ha creato
  - il codice di markup riferisce un **sistema di fonti tipografiche**, che a sua volta definisce una rappresentazione *intelligibile* del documento in forma *grafica*, mediante caratteri stampabili a video o su carta
- **NOTA BENE**: Tali forme di rappresentazione sono gestite diversamente da computer e sistemi operativi diversi

## Esempi

- Documenti Word
- Presentazioni Powerpoint
- Archivi PDF
- E-book
- File HTML
- Sito Web
- ...

## Nomi di documenti

- Nomi all'interno di un computer
- URL: nomi su Web
- URI: identificatori unici
  
- Metadati

7 Struttura globale di un documento HTML

Titoletto primo livello

Sommario

Struttura di un documento

Titoletto secondo livello

Elenco numerato

Sommario

1. Introduzione alla struttura di un documento HTML.
2. Informazioni sulla versione di HTML.
3. L'elemento `<html>`.
4. L'intestazione del documento
  1. L'elemento `<head>`
  2. L'elemento `<title>`
  3. L'attributo `<title>`
  4. I metadati
    - Specificare i metadati
    - L'elemento `<meta>`
    - Profili dei metadati
5. Il corpo del documento
  1. L'elemento `<body>`
  2. Identificatori di elemento: gli attributi `<id>` e `<class>`
  3. Elementi a livello di blocco e in riga
  4. Raggruppare gli elementi: gli elementi `<div>` e `<span>`
  5. Intestazioni: gli elementi `<h1>`, `<h2>`, `<h3>`, `<h4>`, `<h5>`, `<h6>`
  6. L'elemento `<address>`

7.1 Introduzione alla struttura di un documento HTML

Un documento HTML si compone di tre parti:

1. una riga contenente informazioni sulla versione di HTML;
2. una sezione esplicita di intestazione (delimitata dall'elemento `<head>`);
3. un corpo, che contiene il contenuto effettivo del documento. Il corpo può essere implementato per mezzo dell'elemento `<body>` o dell'elemento `<frameset>`.

Dello spazio bianco (spazi, a capo, tabulazioni e commenti) può comparire prima o dopo ciascuna sezione. Le sezioni 2 e 3 dovrebbero essere delimitate dall'elemento `<html>`.

Ecco qui un esempio di un semplice documento HTML:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<HTML>
<HEAD>
<TITLE>Il mio primo documento HTML</TITLE>
```

## Struttura di un documento

- Descrizione **implicita**
- Descrizione **esplicita**
- Una descrizione implicita si ottiene definendo un **tipo** di documento (Es. *libro* oppure *articolo* oppure *lettera* oppure *spartito* oppure...)
- Una descrizione esplicita di struttura si ottiene con un *linguaggio di markup* (es. XML)
- In ogni caso esiste una **grammatica formale** del documento

## Grammatica formale

- Una grammatica formale è un sistema di regole che descrivono un linguaggio
- Es.
  - Libro → Capitolo | Capitolo Libro
  - Capitolo → Paragrafo | Paragrafo Capitolo
  - Paragrafo → Frase "." | Frase Paragrafo
  - Frase → Parola | Parola Frase

## Esempi di strutture

<p>Libro:</p> <ul style="list-style-type: none"> <li>- Capitolo           <ul style="list-style-type: none"> <li>• Paragrafo               <ul style="list-style-type: none"> <li>- Frase. Frase. Frase.</li> </ul> </li> <li>• Paragrafo               <ul style="list-style-type: none"> <li>- Frase.</li> </ul> </li> </ul> </li> <li>- Capitolo           <ul style="list-style-type: none"> <li>• Paragrafo               <ul style="list-style-type: none"> <li>- Frase. Frase.</li> </ul> </li> <li>• Paragrafo               <ul style="list-style-type: none"> <li>- Frase. Frase.</li> </ul> </li> </ul> </li> </ul> <p>Struttura corretta</p>	<p>Libro:</p> <ul style="list-style-type: none"> <li>- Capitolo           <ul style="list-style-type: none"> <li>• Frase               <ul style="list-style-type: none"> <li>- Paragrafo</li> </ul> </li> <li>• Paragrafo               <ul style="list-style-type: none"> <li>- Frase.</li> </ul> </li> </ul> </li> <li>- Capitolo           <ul style="list-style-type: none"> <li>• Capitolo               <ul style="list-style-type: none"> <li>- Frase. Frase.</li> </ul> </li> <li>• Paragrafo               <ul style="list-style-type: none"> <li>- Frase. Frase.</li> </ul> </li> </ul> </li> </ul> <p>Struttura scorretta</p>
--	---

## Esempi di comportamenti

- Transizioni
- Animazioni
- Campi interattivi
- Accesso controllato
- Anteprima
- Formato di stampa
- Layout

Comportamenti **interni**

Comportamenti **esterni**

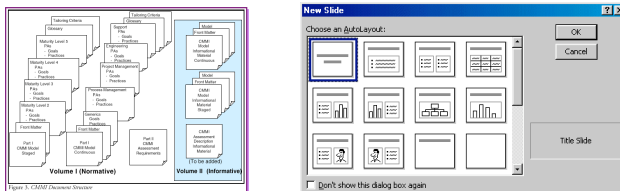
## Altro layout

- Transizioni
- Animazioni
- Campi interattivi
- Accesso controllato
- Anteprima
- Formato di stampa
- Layout

Comportamenti **interni**  
Comportamenti **esterni**

## Struttura e layout

- Non bisogna confondere struttura e layout
- La struttura è una proprietà interna, di solito associata al tipo e al contenuto del documento
- Il layout è un comportamento esterno, dipendente dal dispositivo di visualizzazione



## Dalla struttura al layout

```

<math>
<msqrt>
<mrow>
<msup>
<mi>b</mi>
<mn>2</mn>
</msup>
<mo>-</mo>
<mrow>
<mn>4</mn>
<mi>a</mi>
<mi>c</mi>
</mrow>
</mrow>
</msqrt>
</math>

```

```

<math>
<msubsup>
<mi>A</mi>
<mi>k</mi>
<mn>1</mn>
</msubsup>
</math>

```

$$\sqrt{b^2 - 4ac}$$

## Relazioni con altri documenti

- Contenimento
- Inclusione
- Associazione ipertestuale
- Ancora

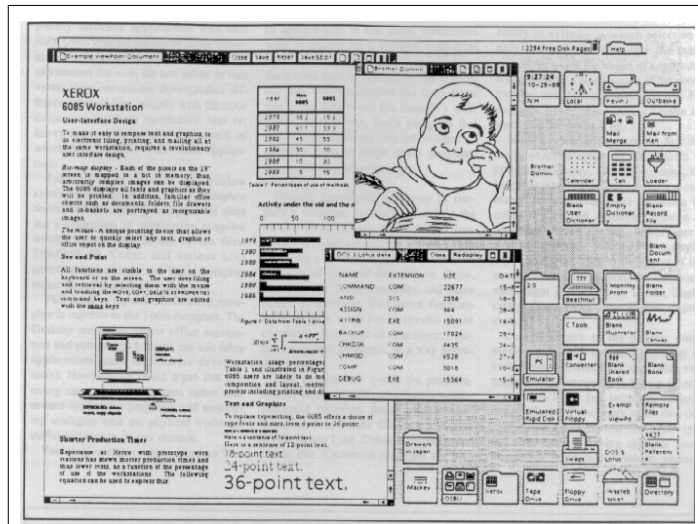
## Metafore

- Documento digitale
- Scrivania
- Iper testo

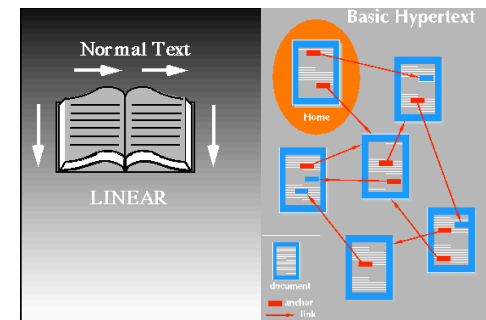
La metafora (dal greco *metaphérō*, «io trasporto») è un tropo (una figura retorica che implica un trasferimento di significato), e si ha quando, al termine che normalmente occuperebbe il posto nella frase, se ne sostituisce un altro la cui "essenza" o funzione va a sovrapporsi a quella del termine originario. Il potere comunicativo della metafora è tanto maggiore quanto più i termini di cui è composta sono lontani nel campo semantico.  
Esempi:

L'Amazzonia è il *polmone* del mondo  
Voi siete il *sale della terra*

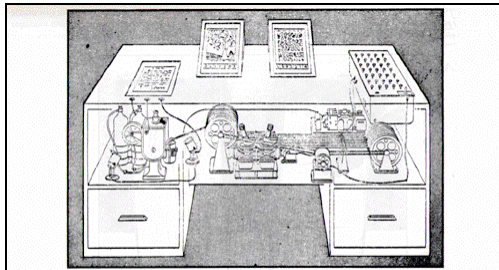
Differisce dalla similitudine per l'assenza di avverbi di paragone o locuzioni avverbiali ("come")."



## Iper testo



## Memex di V. Bush



Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE* 19(11), p.123).

## Memex

“Un memex è un dispositivo in cui una persona memorizza tutti i suoi libri, dischi, e messaggi, e che è meccanizzato in modo che possa essere consultato con grande rapidità e flessibilità”

- Ricerca su tutto il testo, note audio e testuali, link

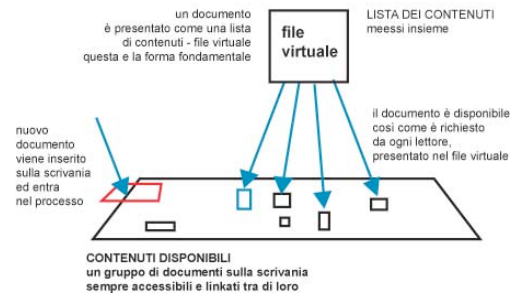
Vannevar Bush, *As We May Think*, 1945



## Xanadu di Nelson

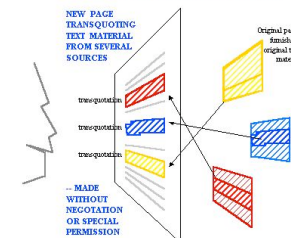


### IL PRINCIPIO DI XANADU



## Transpublishing in Xanadu

### TRANSPUBLISHING: A Basically Simple Concept— Republishing without Complication







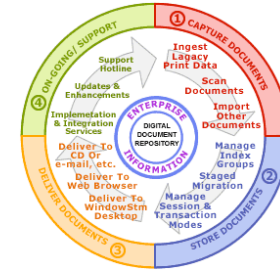
## Esempio di ciclo di vita



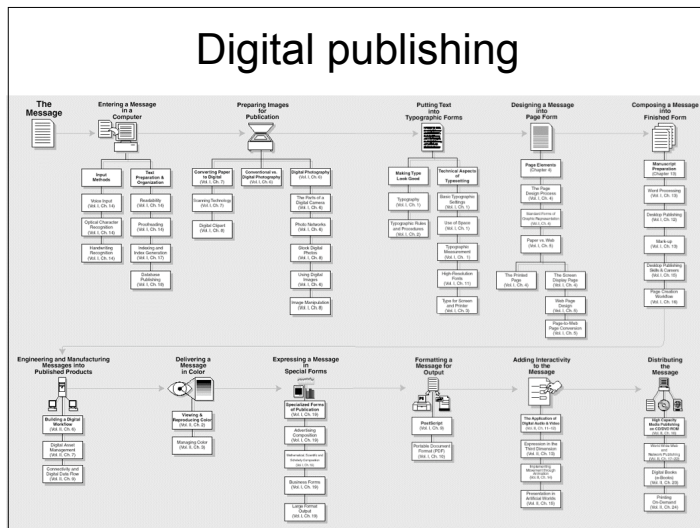
## I documenti nelle organizzazioni

### Strategie d'impiego dei documenti

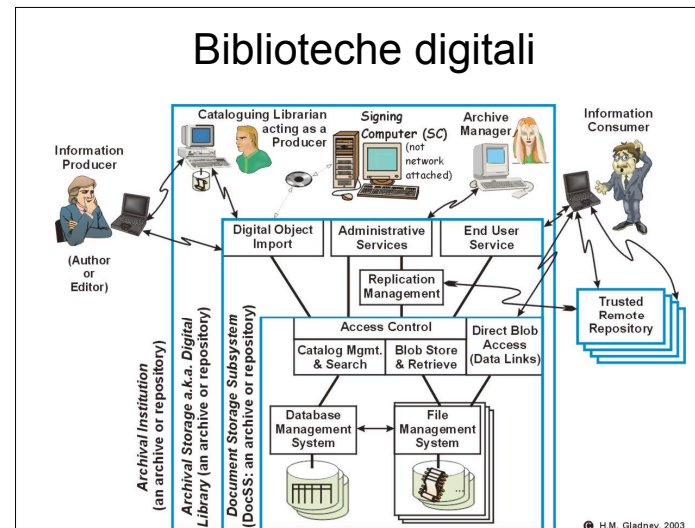
- Strategie di acquisizione
- Strategie di memorizzazione e trasformazione dei contenuti
- Strategie di erogazione e trasmissione
- Architetture informative indipendenti dall'hw



## Digital publishing



## Biblioteche digitali



## Commercio di documenti



Modalità di pagamento   Rinnovo   FAQ - Help   Contatti

Repubblica e gli altri quotidiani extra

# la Repubblica extra

Il quotidiano **La Repubblica** è disponibile anche in **versione digitale**.  
Un servizio utilissimo e pratico per consultare il giornale direttamente sul monitor del proprio computer in formato testo o pdf ed eventualmente stamparlo, per ricercare vecchi articoli grazie all'archivio dell'edizione nazionale degli ultimi 12 mesi o per rileggere le ultime 15 copie del giornale.

La Repubblica extra è la versione digitale del quotidiano venduto ogni giorno in edicola con approfondimenti, notizie sportive, aggiornamenti su tutti gli avvenimenti, tanti servizi utili e la possibilità di leggere comodamente le edizioni locali di Bari, Bologna, Firenze, Genova, Milano, Napoli, Palermo, Roma e Torino.

Consulta subito una copia della Repubblica extra per scoprire tutti i vantaggi di sfogliare e leggere il giornale direttamente sul monitor del computer.

CONSULTA UNA COPIA DEL QUOTIDIANO   SCOPRI I PREZZI E ABBONATI   ←

PACCHETTO PER LE AZIENDE

Contatti

## Rivoluzioni tecnologiche

- Scrittura — 5000 a.C.
- Papiri — 1300 – 800 a.C.
- Stampa — 1455
- Documenti digitali — 1950

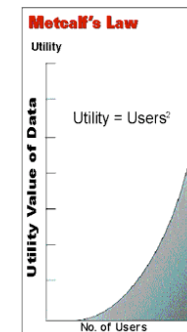
## La terza rivoluzione



- Ridefinì una delle industrie più fiorenti d'Europa
- Aumentò la produttività di migliaia di volte
- Tagliò il costo della riproduzione dei documenti di almeno 1000 volte

## La quarta rivoluzione

- Le tecnologie ICT abbattano i propri costi anno dopo anno
- Struttura esponenziale del valore (Legge di Metcalf): "Il valore di una rete di comunicazione aumenta col quadrato della dimensione della rete stessa"



## What you see is what you get

- Alto
- STAR
- Macintosh



## Macintosh

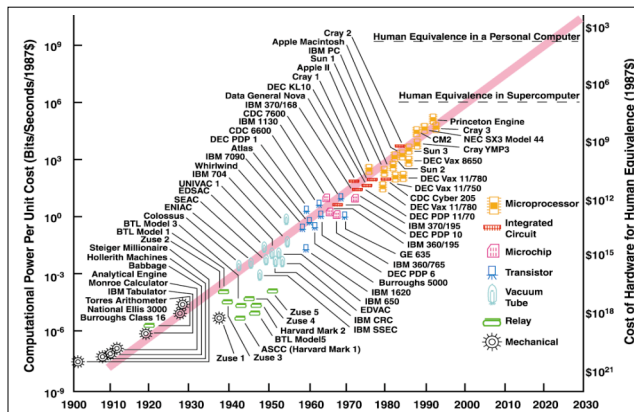
- Macintosh venne creato nel 1984
- Nasce il Desktop Publishing



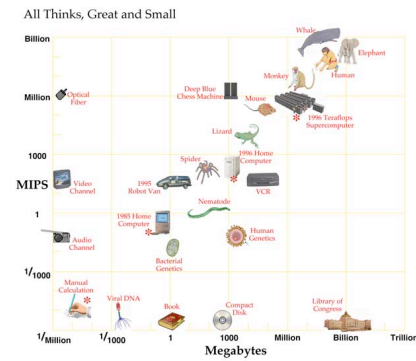
Clip del 1984

Clip del 2004

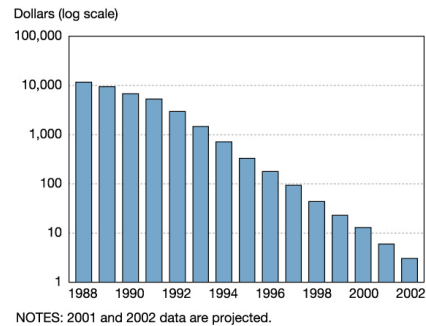
## Legge di Moore



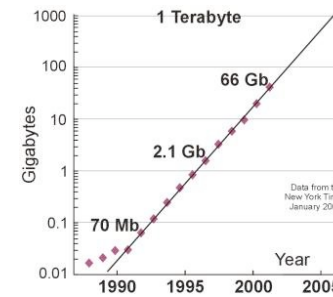
## Legge di Moore



## Costo per gigabyte



## Quanto spazio disco con 200\$?



## A che serve un terabyte?

Cosa si memorizza in un terabyte (= 1000 gigabyte)?

Film	4 Gigabyte ciascuno	250 film
Television-quality video	2 Gigabyte/ora	500 ore
CD musicale	560 Megabyte/disco	1800 ore
Radiografie mediche	10 Megabyte ciascuna	100,000 radiografie
Immagine colore da scanner	1 Megabyte ciascuna	1,000,000 immagini
Documenti B/W da scanner	50,000 bytes/pagina	20,000,000 pagine
Pagine testo ASCII	3300 bytes/pagina	300,000,000 pagine

## Internet Archive [www.archive.org](http://www.archive.org)

Memorizza dal 1996 le pagine del Web, anche dopo che vengono eliminate dalla rete

- Un romanzo 1 MB
- *Encyclopaedia Britannica* (2,619 pagine) 1 GB
- L'antica Biblioteca di Alessandria (400,000 papiri) 800 GB
- Una biblioteca pubblica (300,000 libri) 3 TB
- Un video store (5,000 video, 1 GB/h) 8 TB
- Una stazione radio (10,000 LP/CDs, o 15,000 h, 535MB/h) 8 TB
- Intero Internet Archive al 3/2000 14 TB
- Library of Congress (20 milioni di libri) 20 TB
- Intero Internet Archive al 3/2001 43+ TB
- Intero Internet Archive al 2005 1000+ TB
- Cresce di circa 20TB al mese

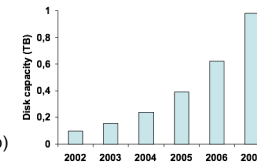
## Progetto MyLifeBits (Gordon Bell)

- Libri scritti (e letti, ove possibile)
- Documenti personali (lettere, conti, avvisi, articoli, note, email...)
- Foto
- Collezione CD
- Dipinti e poster
- Filmati casalinghi
- Tutti i file sul PC
- Chiamate telefoniche
- Trasmissioni radio e TV ascoltate
- Pagine web visitate
- ...

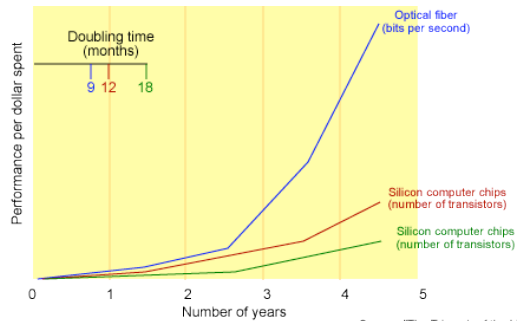


## Memorizzare la vita

- Conservando:
  - 100 messaggi email al giorno (5KB ciascuno)
  - 100 pagine web al giorno (50KB ciascuno)
  - 5 pagine da scanner al giorno (100KB ciascuno)
  - 1 libro ogni 10 giorni (1 MB ciascuno)
  - 10 foto al giorno (400 KB JPEG ciascuno)
  - 8 ore al giorno di suoni - e.g. telefonate, note vocali, registrazioni di incontri (8 Kb/s)
  - 1 nuovo CD musicale ogni 10 giorni (45 min each at 128 Kb/s)
- Occorrono 5 anni per riempire un disco da 80 GB
- Per riempire un disco da un terabyte occorrono più di 60 anni

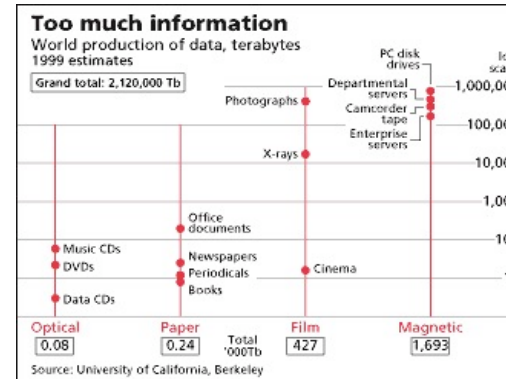


## Il trionfo della luce



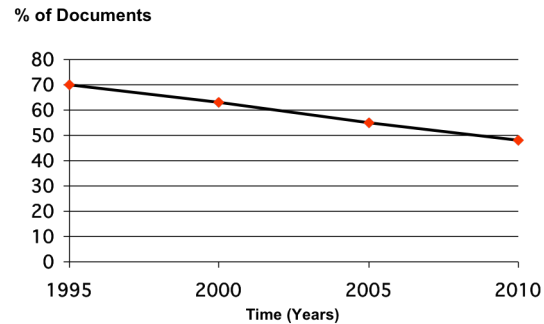
Source: "The Triumph of the Light," Scientific American, January 2001

## Aumenta la carta?



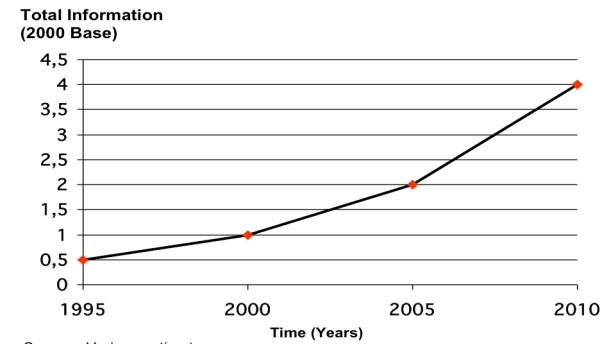
Source: University of California, Berkeley

## Informazioni su carta



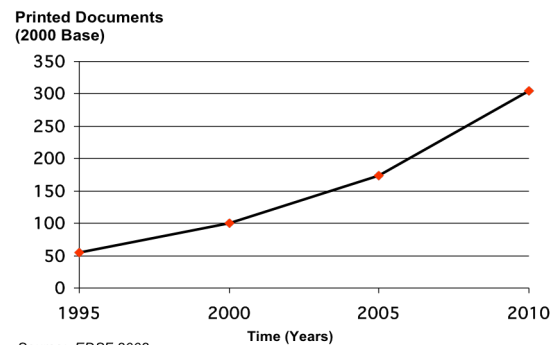
Source: RIT Analysis 2000

## ...ma le informazioni crescono



Source: Various estimates

## ...dunque cresce anche la carta



Source: EDSF 2002

## L'età digitale: Internet

- Oltre un miliardo di persone la usano
  - Il Nord America al 70%
  - Al 35% in EMEA (Europe, MiddleEast, Asia)
  - Digital divide

## L'impatto sul commercio

- 55% dei siti che fanno commercio elettronico hanno profitti
- Aziende: 70% delle grandi, 40% delle piccole usano il commercio elettronico
- Fatturato complessivo mondiale annuo verso 100 miliardi\$, di cui 23% computer, 13% **libri**
- Per altro, l'industria musicale nel 2002 ha perso il 10% del fatturato per colpa di Internet

## Aumenta la carta stampata, ma l'industria editoriale decade

- Numero di imprese editoriali in calo - 7000 in meno nel decennio dei '90 (Trendwatch)
- Gli introiti editoriali complessivi sono in calo anche se aumentano i profitti - RIT
- \$36 miliardi di perdita potenziale dell'industria dovuta alla tecnologia digitale
- 70% degli introiti prepress persi nel periodo 1992-2000
- "La crescita dell'industria editoriale non è più coerente con la crescita del PNL" *NAPL - 2003*

## Trend contemporanei

- Aumento dei supporti didattici basati su nuove tecnologie
- Declino della posta cartacea negli U.S.A. e nelle altre nazioni sviluppate
- La percentuale di pubblicità su carta in declino rispetto al totale

## Trend contemporanei

- Declino delle attività basate su carta: giornali, libri, riviste
- Le nuove generazioni usano meglio e più diffusamente le tecnologie informatiche
- Miglioramenti tecnologici
  - Banda sempre più larga
  - Migliore mobilità dei dispositivi
  - Schermi più raffinati e leggibili
  - Digitalizzazione completa delle risorse editoriali



## Riferimenti

- W.Kasdorf, *The Columbia Guide to Digital Publishing*, Columbia UP, 2003
- V.Bush, *As We May Think*, *Atlantic Monthly*, 1945
- T.Nelson, *Literary Machines*, 1981
- G.Bell e altri, *MyLifeBits: Fulfilling the Memex vision*, [research.microsoft.com/~jgemell/pubs/MyLifeBitsMM02.pdf](http://research.microsoft.com/~jgemell/pubs/MyLifeBitsMM02.pdf)

## Siti

[seminars.seyboldreports.com](http://seminars.seyboldreports.com)

Sito con documenti dedicati all'industria dei documenti digitali

- [www.fred.net/tds/longrange.html](http://www.fred.net/tds/longrange.html)
- [www.press.umich.edu/jep/](http://www.press.umich.edu/jep/)
- [www.publishing-industry.net/](http://www.publishing-industry.net/)
- [www.colormatters.com/](http://www.colormatters.com/)

## Domande?