INFO116

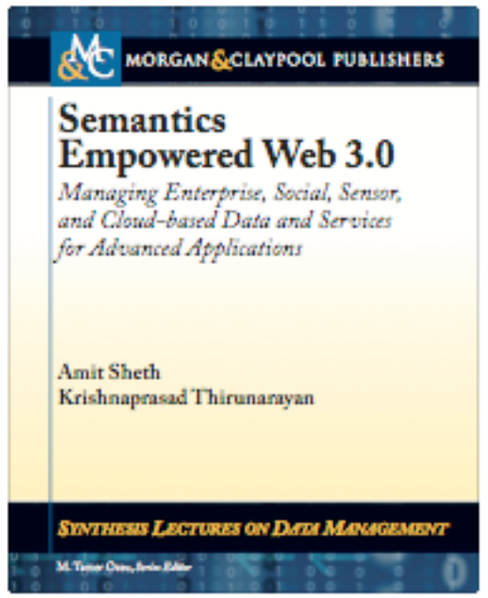
Role of Semantics and Metadata

WELCOME

- Course
  - Lectures (16-20 hours)
  - Labs (16-20 hours)
- Assessment
  - 3 hours Written Exam (60%)
  - 1 Group Assignment (40%)

READINGS

- Text: http://www.morganclaypool.com/doi/abs/10.2200/S00455ED1V01Y201211DTM031
- Additional notes on miside
- Lab notes



METADATA

- What is metadata?

drwxr-xr-x 4 cveres staff 136 20 Aug 15:34 ./

drwxr-xr-x 15 cveres staff 510 21 Aug 08:52 ../

-rw-r--r--@ 1 cveres staff 301 20 Aug 15:36 classes.n3

-rw-r--r--@ 1 cveres staff 722 20 Aug 15:26 ex1.n3

SEMANTICS

- Semantics = “Meaning”
  - What is “meaning”?
  - “Everybody loves somebody.”
  - Montague grammar, Prototype theory, Model theoretic semantics, Lexical and conceptual semantics, semantics of programming languages, memory for meaning, ...
- Meaning (for us) = Metadata++

WHY DO WE NEED SEMANTICS?

- Web 3.0 has become a repository of an ever growing variety of Web resources
  - enterprise data and services
  - social networks
  - cloud computing
  - mobile
  - “Internet of things”

CHALLENGES

- "Big Data"
  - heterogeneity (variety)
  - scale (volume)
  - continuous changes (velocity)

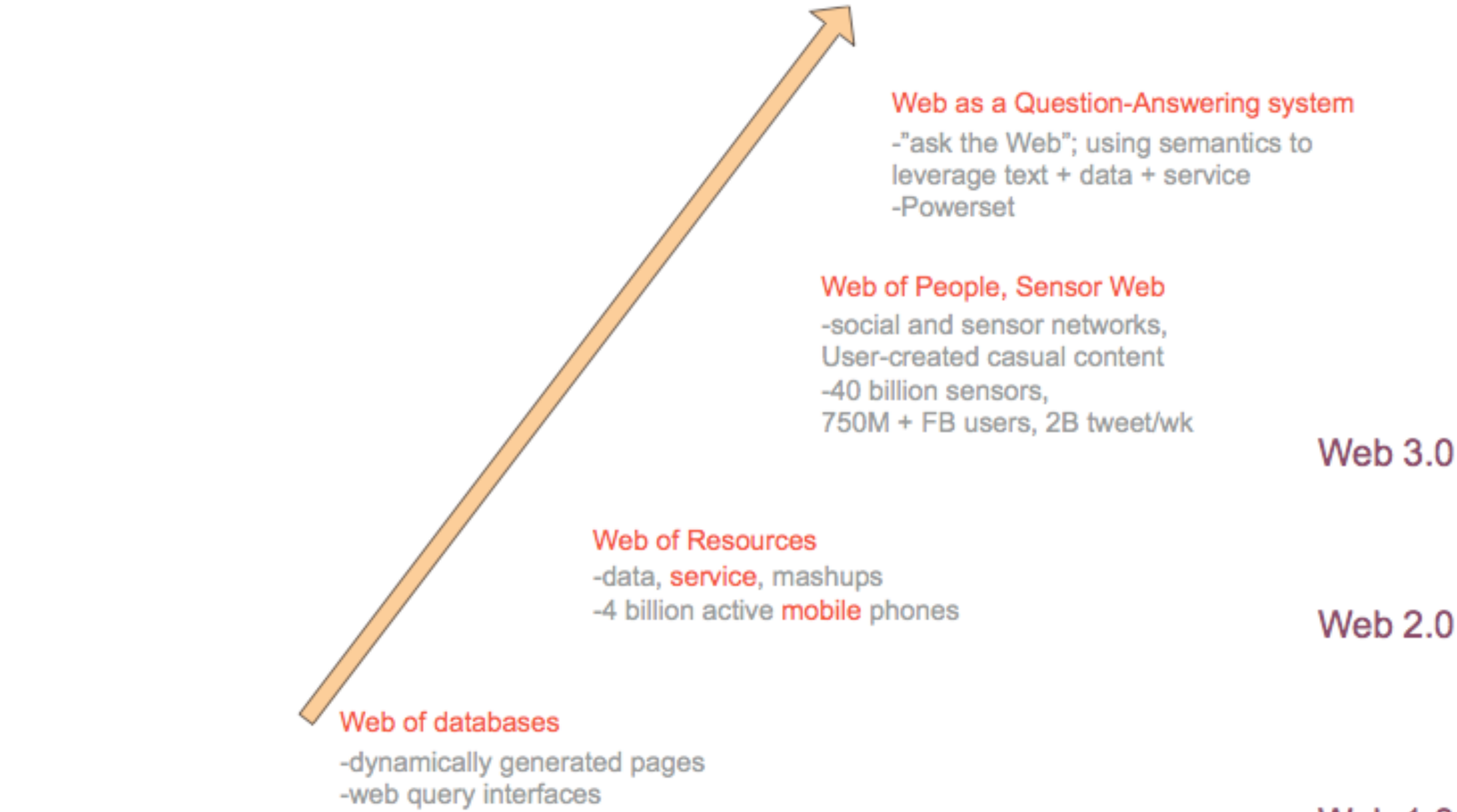
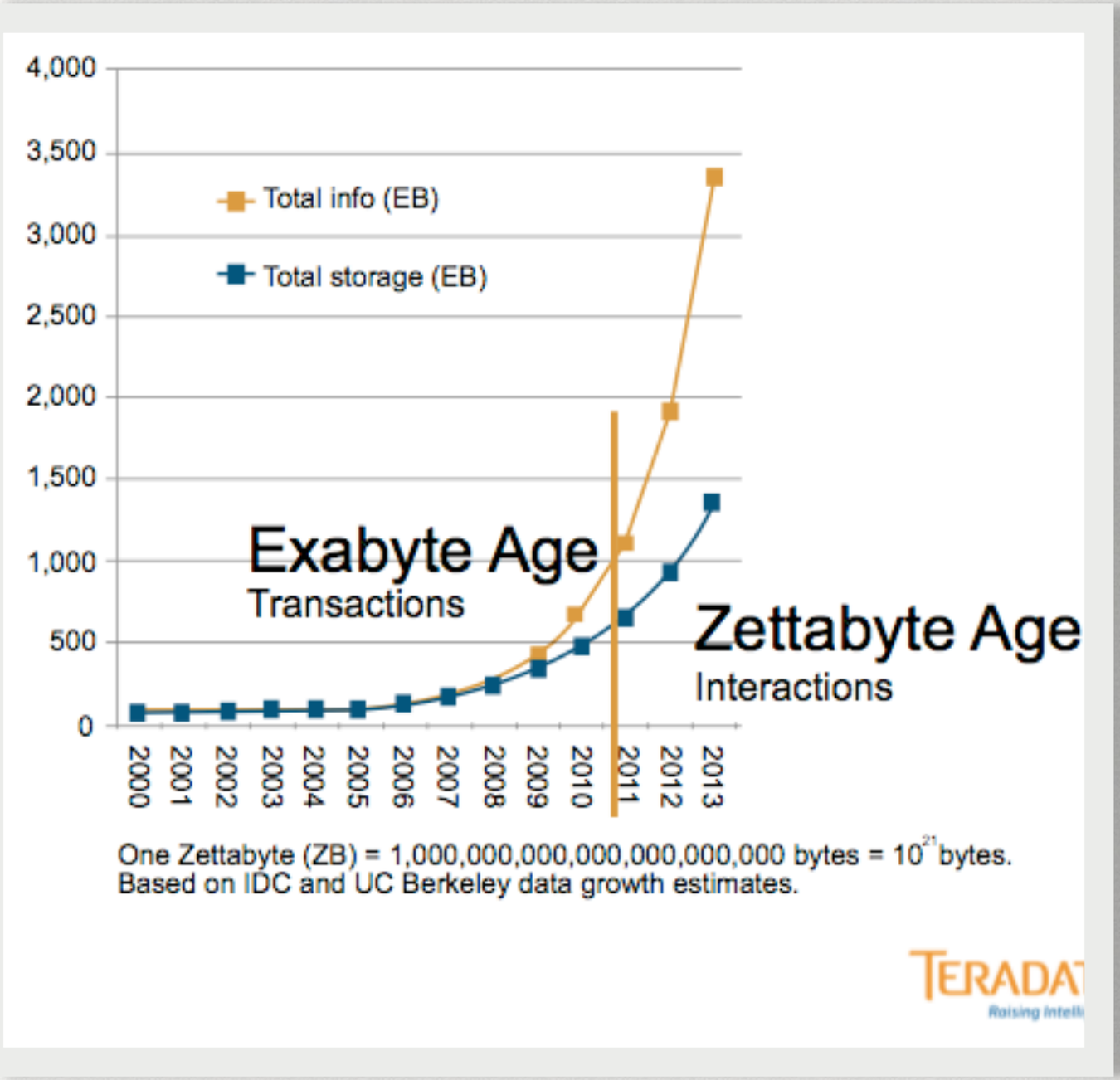


Figure 1.1: Evolution of Web.

BIG DATA

- A Boeing jet generates 10 terabytes of information per engine every 30 minutes of flight, according to Stephen Brobst, the CTO of Teradata. So for a single six-hour, cross-country flight from New York to Los Angeles on a twin-engine Boeing 737 — the plane used by many carriers on this route — the total amount of data generated would be a massive 240 terabytes of data. There are about 28,537 commercial flights in the sky in the United States on any given day.



WHAT TO DO?

- We can no longer store all the data that are generated
- But we want to
  - search, browse, integrate, mine, ...
  - ultimately use them to gain insight, develop situational awareness, discover new knowledge, answer difficult questions, and make decisions.
- Automatically relate, disambiguate, understand, and abstract data

HOW DID WE GET INTO THIS MESS?

- 1980s: client-server paradigm
- For each function, there was a separate server and client program.
- FTP was used to share files
- TELNET was used to access different machines
- EMAIL was used for private communication
- USENET was used for public exchanges.

WEB IS BORN

- 30 April 1993

950430

ORGANISATION EUROPEENNE POUR LA RECHERCHE NUCLEAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

STATEMENT CONCERNING CERN W3 SOFTWARE RELEASE INTO PUBLIC DOMAIN

TO WHOM IT MAY CONCERN

Introduction

The World Wide Web, hereafter referred to as W3, is a global computer networked information system.

The W3 project provides a collaborative information system independent of hardware and software platform, and physical location. The project spans technical design notes, documentation, news, discussion, educational material, personal notes, publicity, bulletin boards, live status information and numerical data as a uniform continuum, seamlessly intergated with similar information in other disciplines.

The information is presented to the user as a web of interlinked documents .

Acces to information through W3 is:

- via a hypertext model;
- network based, world wide;
- information format independent;
- highly platform/operating system independent;
- scalable from local notes to distributed data bases.

Webs can be independent, subsets or supersets of each other. They can be local, regional or worldwide. The documents available on a web may reside on any computer supported by that web.

Declaration

The following CERN software is hereby put into the public domain:

- W 3 basic ("line-mode") client
- W 3 basic server
- W 3 library of common code.

CERN's intention in this is to further compatibility, common practices, and standards in networking and computer supported collaboration. This does not constitute a precedent to be applied to any other CERN copyright software.

CERN relinquishes all intellectual property rights to this code, both source and binary form and permission is granted for anyone to use, duplicate, modify and redistribute it.

CERN provides absolutely NO WARRANTY OF ANY KIND with respect to this software. The entire risk as to the quality and performance of this software is with the user. IN NO EVENT WILL CERN BE LIABLE TO ANYONE FOR ANY DAMAGES ARISING OUT THE USE OF THIS SOFTWARE, INCLUDING, WITHOUT LIMITATION, DAMAGES RESULTING FROM LOST DATA OR LOST PROFITS, OR FOR ANY SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES.

Geneva, 30 April 1993

W. Hoogland
Director of Research

H. Weber
Director of Administration

opie certifiée conforme

ait à Genève le 03-05-93

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
DIVISION DES FINANCES
CERN

Tim Berners-Lee with Nicola Pellow, the writer of the line-mode browser for the World Wide Web next to the NeXT computer in 1991.

World Wide Web

The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents.

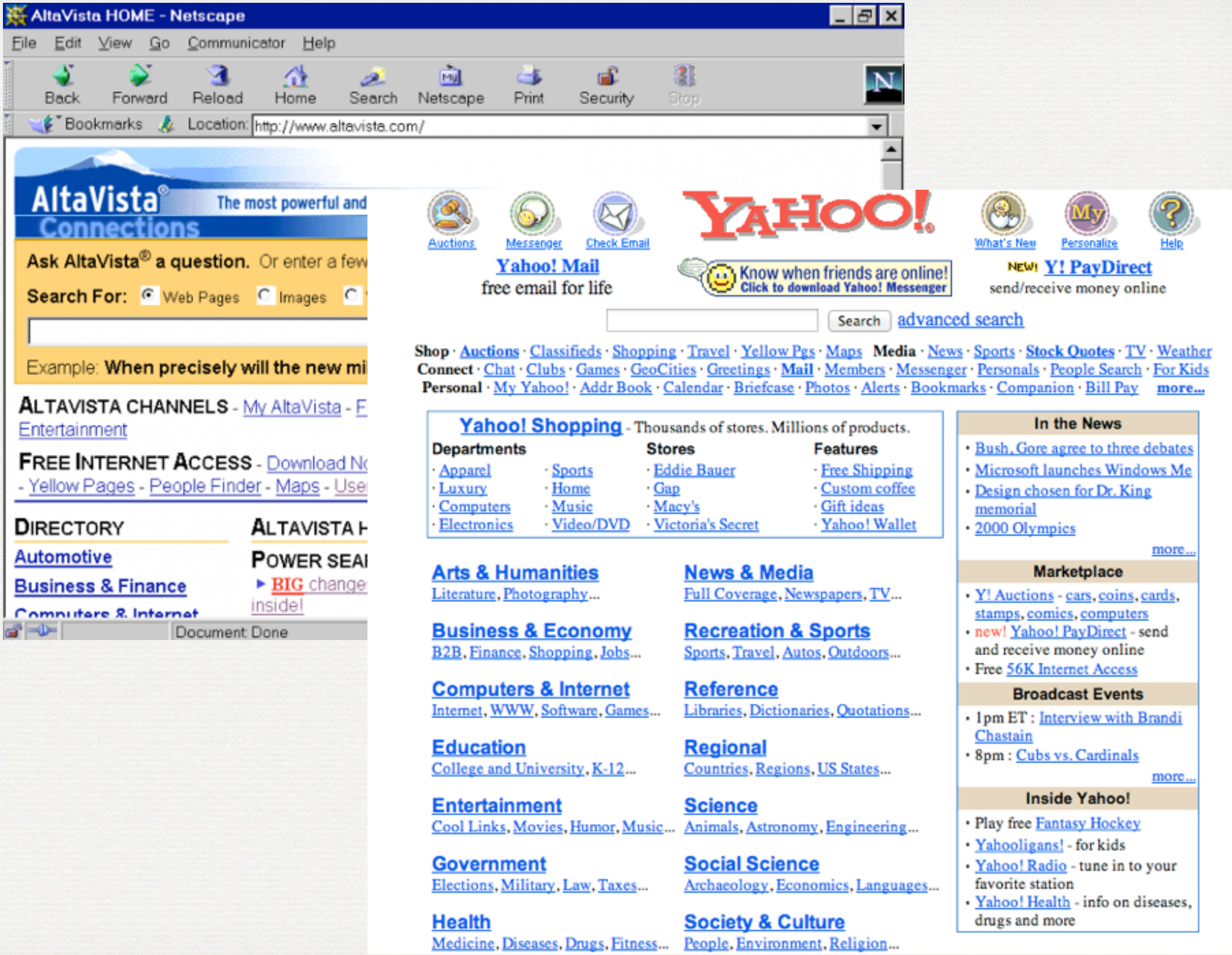
Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists , Policy , November's W3 news , Frequently Asked Questions .

What's out there?
Pointers to the world's online information, subjects , W3 servers, etc.
Help
on the browser you are using
Software Products
A list of W3 project components and their current state. (e.g. Line Mode ,X11 Viola , NeXTStep , Servers , Tools , Mail robot , Library)
Technical
Details of protocols, formats, program internals etc
Bibliography
Paper documentation on W3 and references.
People
A list of some people involved in the project.
History
A summary of the history of the project.
How can I help ?
If you would like to support the web..
Getting code
Getting the code by anonymous FTP , etc.

- Uniform Resource Locator (URL)
  - communication protocol [http, telnet, ftp, mail, etc.]
  - machine [IP address]
  - full-path address [in a directory]
- HyperText Transmission Protocol (HTTP) was developed for text-based request-response communication
- HyperText Markup Language (HTML), an application/instance of SGML

SEARCH

- As web grew, people needed some way of finding stuff
  - keyword search
    - pure keyword
    - TF/IDF
    - links (page rank) (implicit metadata?)
  - category systems (semantics!)


AltaVista HOME - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location: http://www.altavista.com/
AltaVista® The most powerful and
Connections
Ask AltaVista® a question. Or enter a few
Search For: Web Pages Images
Example: When precisely will the new mi
ALTAVISTA CHANNELS - My AltaVista - F
Entertainment
FREE INTERNET ACCESS - Download No
- Yellow Pages - People Finder - Maps - Use
DIRECTORY
Automotive
Business & Finance
Computers & Internet
ALTAVISTA H
POWER SEAI
BIG change
inside!
Document Done
Auctions Messenger Check Email
YAHOO!
What's New Personalize Help
Yahoo! Mail
free email for life
Know when friends are online!
Click to download Yahoo! Messenger
NEW! Y! PayDirect
send/receive money online
Search advanced search
Shop · Auctions · Classifieds · Shopping · Travel · Yellow Pgs · Maps Media · News · Sports · Stock Quotes · TV · Weather
Connect · Chat · Clubs · Games · GeoCities · Greetings · Mail · Members · Messenger · Personals · People Search · For Kids
Personal · My Yahoo! · Addr Book · Calendar · Briefcase · Photos · Alerts · Bookmarks · Companion · Bill Pay more...
Yahoo! Shopping - Thousands of stores. Millions of products.
Departments
· Apparel · Luxury · Computers · Electronics
· Sports · Home · Music · Video/DVD
Stores
· Eddie Bauer · Gap · Macy's · Victoria's Secret
Features
· Free Shipping · Custom coffee · Gift ideas · Yahoo! Wallet
In the News
· Bush, Gore agree to three debates
· Microsoft launches Windows Me
· Design chosen for Dr. King memorial
· 2000 Olympics
more...
Marketplace
· Y! Auctions - cars, coins, cards, stamps, comics, computers
· new! Yahoo! PayDirect - send and receive money online
· Free 56K Internet Access
Broadcast Events
· 1pm ET : Interview with Brandi Chastain
· 8pm : Cubs vs. Cardinals
more...
Inside Yahoo!
· Play free Fantasy Hockey
· Yahooligans! - for kids
· Yahoo! Radio - tune in to your favorite station
· Yahoo! Health - info on diseases, drugs and more
Arts & Humanities
Literature, Photography...
Business & Economy
B2B, Finance, Shopping, Jobs...
Computers & Internet
Internet, WWW, Software, Games...
Education
College and University, K-12...
Entertainment
Cool Links, Movies, Humor, Music...
Government
Elections, Military, Law, Taxes...
Health
Medicine, Diseases, Drugs, Fitness...
News & Media
Full Coverage, Newspapers, TV...
Recreation & Sports
Sports, Travel, Autos, Outdoors...
Reference
Libraries, Dictionaries, Quotations...
Regional
Countries, Regions, US States...
Science
Animals, Astronomy, Engineering...
Social Science
Archaeology, Economics, Languages...
Society & Culture
People, Environment, Religion...

dmoz open directory project

In partnership with AOL Search.

about dmoz | dmoz blog | suggest URL | help | link | editor login

Search *advanced*

Arts
Movies, Television, Music...

Business
Jobs, Real Estate, Investing...

Computers
Internet, Software, Hardware...

Games
Video Games, RPGs, Gambling...

Health
Fitness, Medicine, Alternative...

Home
Family, Consumers, Cooking...

Kids and Teens
Arts, School Time, Teen Life...

News
Media, Newspapers, Weather...

Recreation
Travel, Food, Outdoors, Humor...

Reference
Maps, Education, Libraries...

Regional
US, Canada, UK, Europe...

Science
Biology, Psychology, Physics...

Shopping
Clothing, Food, Gifts...

Society
People, Religion, Issues...

Sports
Baseball, Soccer, Basketball...

World
Català, Dansk, Deutsch, Español, Français, Italiano, 日本語, Nederlands, Polski, Русский, Svenska...

Become an Editor Help build the largest human-edited directory of the web

Copyright © 2013 Netscape

5,223,457 sites - 98,564 editors - over 1,019,020 categories

excite.

My Email | My Money | My Scores | My Weather | My Videos | My TV | My Movies

Hi Csaba! Settings • Content • Layout • Colors & Themes • Sign Out My Links: show My Pages: hide Lite • Email • Help Min

Advertisement

My Page (edit) add / delete pages

Web Images News Video

Search the Web: Search Got Speedbar?

Yellow Pages | White Pages

Best Fares

Expedia

Flight Packages
Hotel Cruise
Car Activities

Shop with confidence:
Best price guaranteed!

Quote of the Day

The one resolution, which was in my mind long before it took the form of a resolution, is the key-note of my life. It is this, always to regard as mere impertinences of fate the handicaps which were placed upon my life almost at the

Explore Excite

Air/Travel	Entertainment	Money
Autos	Family	Movies
Beauty	Fashion	News
Careers	Games	Real Estate
Casino	Gifts	Recipes
Celebrities	Health	Recreation
Computers	Holidays	Shopping
Crafts	Home & Garden	Software
Credit Score	How To Video	Sports
Dating	Lifestyle	Technology
eBay®	Local	Today
		Video NEW

Shop Books, Air Tickets, Classifieds...

Best Fares

Expedia

Flight Packages
Hotel Cruise
Car Activities

Shop with confidence:
Best price guaranteed!

2003. WEB 2.0

- read / write web
- users generated content
- interact and collaborate with each other in a social media
  - social networking sites, blogs, wikis, video sharing sites, folksonomies,

WEB 2.0 TECHNOLOGIES

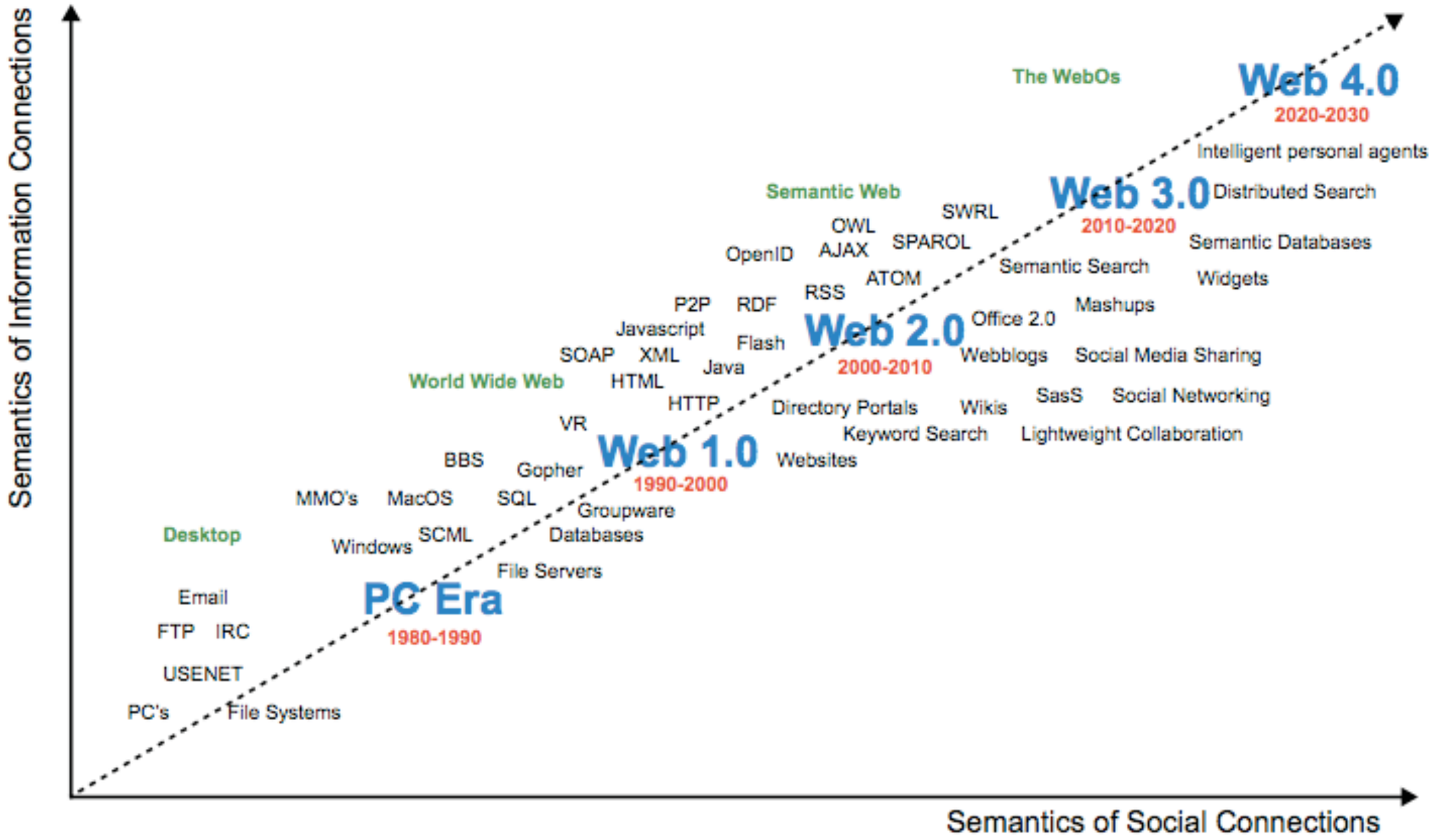
- Emerging new standards and technologies
- XML
- AJAX (Asynchronous JavaScript and XML that enables Web applications to retrieve data from the server asynchronously without interfering with the existing page)
- JSON (JavaScript Object Notation, lightweight data interchange)
- RSS (Really Simple Syndication that allows subscription to feeds such as News, Events, Sports, etc.)
- P2P (Peer to Peer) content sharing systems

WHAT ABOUT SEMANTICS ?

- Tags (early promise, but limited)
- Clustering, aggregation
- By 2006, there were 100 million websites, 9 billion web pages, and 1 billion global users.
- Something had to be done!

2006. WEB 3.0

- Currently there are approximately 250 million websites and 32 billion web pages
- If we are to have any chance of organizing, integrating and understanding this mass of data, then the data will have to be self describing in some way: we have to know what the data is about, and machines will have to be able to reason about the data.



Source: Radar Networks & Nova Spivack, 2007 - www.radarnetworks.com

A SEMANTIC WEB

- "If HTML and the Web made all the online documents look like one huge book, RDF, schema, and inference languages will make all the data in the world look like one huge database"
  - integrating information about the same concept or object in a different modality and media
- Intelligent processing: objects, relationships, semantic search instead of text and keyword search.
- Knowledge enabled computing
- Abstraction

SOME POSSIBILITIES

- Providing semantics to data involves understanding entities, actions, and relationships the data describe
  - Implicit semantics: the meaning of the data is in the patterns which are not explicitly represented.
  - Formal semantics. Use of an artificial logical language designed for knowledge representation and reasoning.

IMPLICIT SEMANTICS

- Link structure
- Clusters, word co-occurance
- Contextual disambiguation

FORMAL SEMANTICS

- Family of logics
  - FOL
  - Higher order logics
    - modal, deontic, etc.
  - Description logics
- Compositionality