CSE490i Advanced Internet Systems

- **Prof: Oren Etzioni**
  - Lectures, concepts, perspective.
- **TA: Valentin Razmov**
  - Project, course web, machine/environment, detailed questions.
- **Class: TTh 12-1:20pm.**
- **Expectations:**
  - Project (multiple parts, on time!)
  - Reading (papers, web - no formal text)
  - Class participation / development
  - final (possible)
- **Caveat: Life on the cutting edge**

Personal Motivation

- **Research on Intelligent Internet Systems**
  - Since 1991
  - Internet softbot (Discover award finalist '95)
  - Webcrawler by Brian Pinkerton
  - Metacrawler by Eric Selberg & Oren Etzioni
- **Netbot co-founder (sold to Excite '97).**
- **CTO Go2net (merged w/ infospace '00).**
- **Venture partner w/ Madrona Venture Group.**
- **Incredible shortage of software engineers!**
- **Dearth of training**

Syllabus

- **Introduction**
- **Search Engine Technology**
- **Meta-search, query routing**
- **Website Management (XML, DBs)**
- **Personalization, adaptive web sites**
- **Security**
- **Internet Business Models**

Project

- **Build an MP3 portal.**
- **Four Cumulative Parts**
  - Write crawler that extracts content;
  - Add database connectivity/storage.
  - Basic website management (XML, Javascript)
  - Extend site in a clever way (natural-language queries, adaptation, meta-search, …)

Warning

- **490i midway between 4xx class & capstone design**
- **No textbook**
- **Poorly documented, unstable systems**
- **We have never taught it before**
- **Need students to help debug class!**

Input Student Info

- **Name**
- **Email**
- **Background**
  - 444?
  - 451?
  - 461?
  - 473?
  - Other 4xx?
- **Languages**
  - Javascript?
  - Java?
  - Others:
- **What are our plans upon graduation?**

Abbreviated History

Pre-history: the Greeks, Turing, Census, WWII.
1943 First electronic digital computer Harvard Mark I
1960 Ted Nelson proposes Xanadu
1961 Kleinrock paper on packet switching (unlike phone lines, which are circuit switched).
1965 Gordon Moore proposes law
1966 Design of ARPAnet
1968 Doug Engelbart: the first WIMP
1969 First ARPAnet message UCLA -> SRI

History

1970 ARPAnet spans country, has 5 nodes
1971 ARPAnet has 15 nodes
1972 First email programs, FTP spec
1973 Ethernet operation at Xerox PARC
1974 Intel launches 8080; TCP design
1975 Gates/Allen write Basic for Altair 8800
1976 Apple Computer formed by Jobs/Wozniak
1977 111 hosts on ARPAnet
1979 Visicalc

History

1981 Microsoft has 40 employees; IBM PC
1982 Sun formed
1983 ARPAnet uses TCP/IP -> birth of internet
1983 Design of DNS
1984 launch of Macintosh; 1000 hosts on ARPAnet
1985 Symbolic.com first registered domain name
1989 100,000 hosts on Internet
1990 Cisco Systems goes public $288 M
Tim Berners-Lee creates WWW at CERN

History

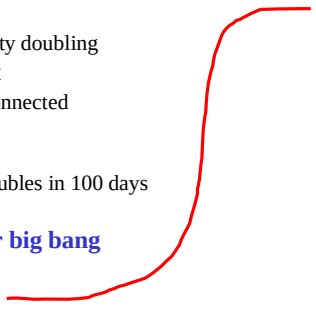
1993 Mosaic developed at UIUC
Web grows by 341,000% in a year
1994 Netscape, Amazon, Archtext formed
1995 Netscape IPO, Windows 95, MetaCrawler
1997 Amazon IPO
2000 Internet "bubble" crashes.
2001 ???

Take Home Messages

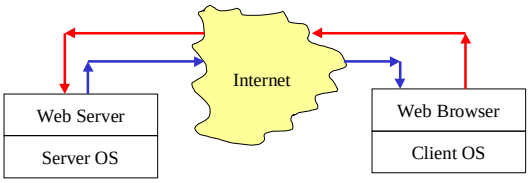
- **Universities, private sector, and DARPA all contributed.**
- **History is discontinuous.**
- **Standards change the world.**
- **Not everything originates in the USA.**
- **What else?**

The Internet (Big and Getting Bigger)

- **Moore's Law**
  - Semiconductor density doubling
- **Internet Equivalent**
  - 1996: 40M people connected
  - 1997: 100M
  - 1998 : Traffic vol doubles in 100 days
- **First moments after big bang**

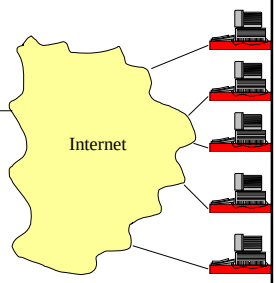


Connecting on the WWW



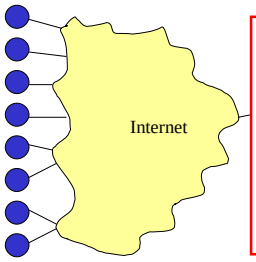
Server-Side View

- **Database-driven content**
- **Lots of Users**
  - **Scalability**
  - **Load balancing**
- **Often implemented with cluster of PCs**
- **24x7 Reliability**
- **Transparent upgrades**



Clients

Client-Side View



Content rendering engine
- Tags, positioning, movement

Scripting language interpreter
- Document object model
- Events
- Programming language itself

Link to custom Java VM
Security access mechanisms
Plugin architecture + plugins

Web Sites

Trade-offs in Client/Server arch.

- **Compute on servers?**
  - + easy to control, debug.
  - – Peak load, reliability, cost.
- **Compute on clients?**
  - – Many different browsers
  - – {Netscape, IE, Lynx, …} × Version × OS
  - – Each supports different tags, languages…
- **Peer-to-peer? Gnutella?**