

# Chapter 1:

## Basic Concepts

Instructor: Dr. Mehmet S. Aktaş

---

# Road map

- World Wide Web
- Internet
- Search Engines
- Web Data Mining
  - Why? What? Where?
- Summary

---

# World Wide Web - WWW

- The Web
  - allows users of one computer to access information stored on another through a world-wide network called the Internet.
- Client-server based implementation
  - Ex: local machine browser, remote machine: Web servers
- Hypertext document based semantic structure
  - Ted Nelson in 1965
  - Hypertext, hyperlinks

---

# World Wide Web - History

## ■ The Web

- ❑ Invented by Tim Berners-Lee at CERN in 1989 (19 years ago!!!).
- ❑ Distributed hypertext system – basic architecture of the Web.
- ❑ Hypertext Transfer Protocol (HTTP), Hypertext Markup Language (HTML), Universal Resource Locator (URL)

## ■ Browsers

- ❑ Mosaic, 1993, M. Andreessen, Univ. of Illinois
- ❑ Netspace, 1994, M. Andreessen and Jim Clark, Netscape Comm.
- ❑ Internet Explorer, 1995, B. Gates, Microsoft

# Internet - History

## ■ ARPANET - Internet

- Advanced Research Projects Agency (ARPA) Network, 1969-1972.
  - Connected computers from 40 different locations.
- TCP/IP (Transmission Control Protocol/Internet Protocol), 1973, Vinton Cerf, Bob Kahn.
- Internet
  - 1982 TCP/IP was widely adopted.
  - Networks were connected using TCP/IP Protocol and formed inter-net.

# Search Engines

- Excite search system
  - 1993, started by six Stanford Univ. students
- Yahoo
  - 1994, started out as a listing of favorite web sites
  - 2004, provided general search capability with Inktomi
- Many Search Engines
  - 1994 - present, Lycos, Infoseek, Alta Vista, Ask Jeeves etc.
- Google
  - 1998 by Sergey Brin and Larry Page at Stanford Univ.
- MSN
  - 2005 by Microsoft, which used others search engines before

---

# Characteristics of the Web

- The Web has huge amount of available data
- The Web has all types of data
- The Web has heterogeneous information
- The Web has a hyperlink structure
- The Web has noisy information
- The Web is about services
- The Web is very dynamic
- The Web is a virtual society

# Web Data Mining

## ■ Web Mining

- aims to discover useful information from the Web hyperlink structure, page content and usage data.
- Web structure mining
  - Discovers knowledge from hyperlinks
- Web content mining
  - Mines useful information from Web page contents
- Web usage mining
  - Mines useful information from Web usage logs