

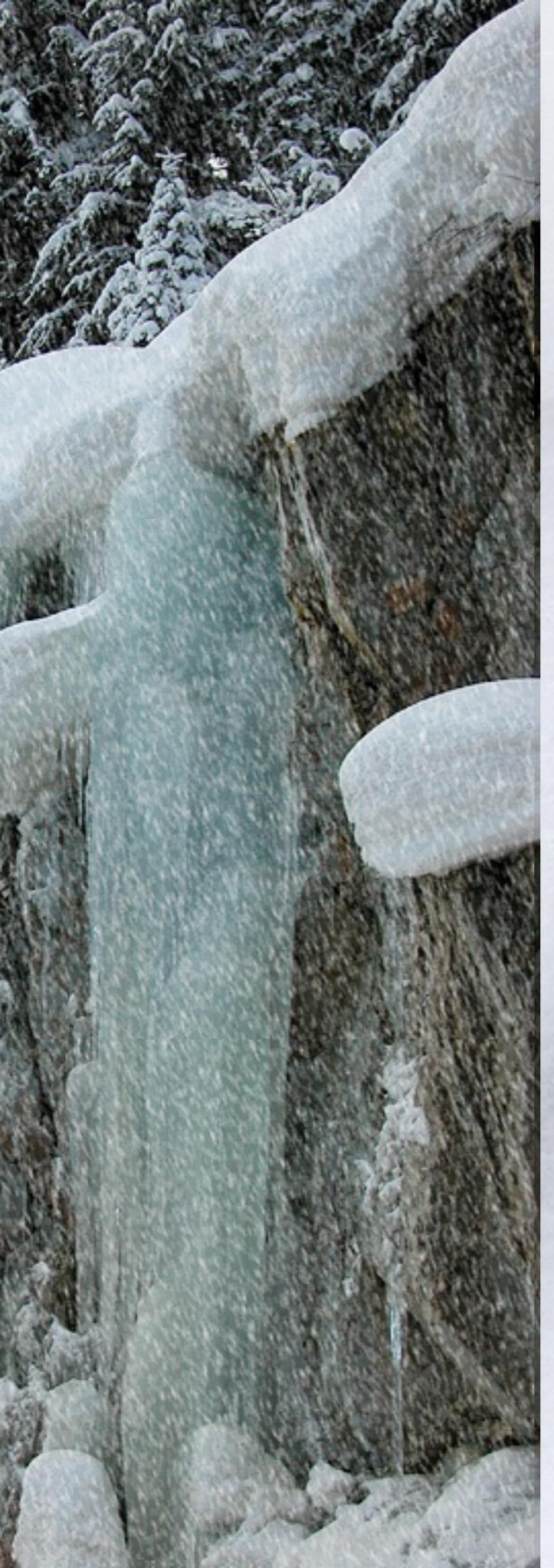
Lecture 1

Internet and World Wide Web



Play

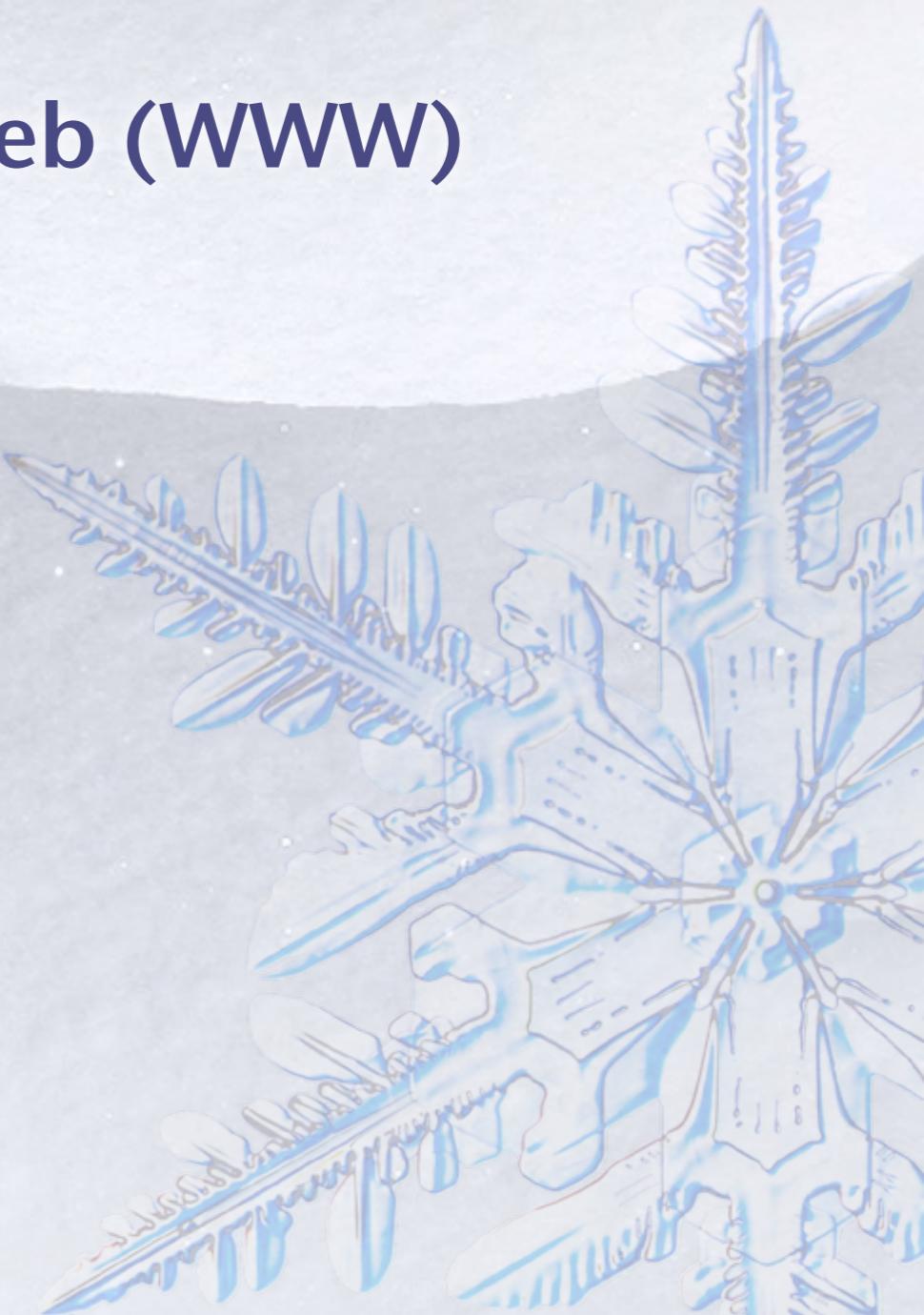




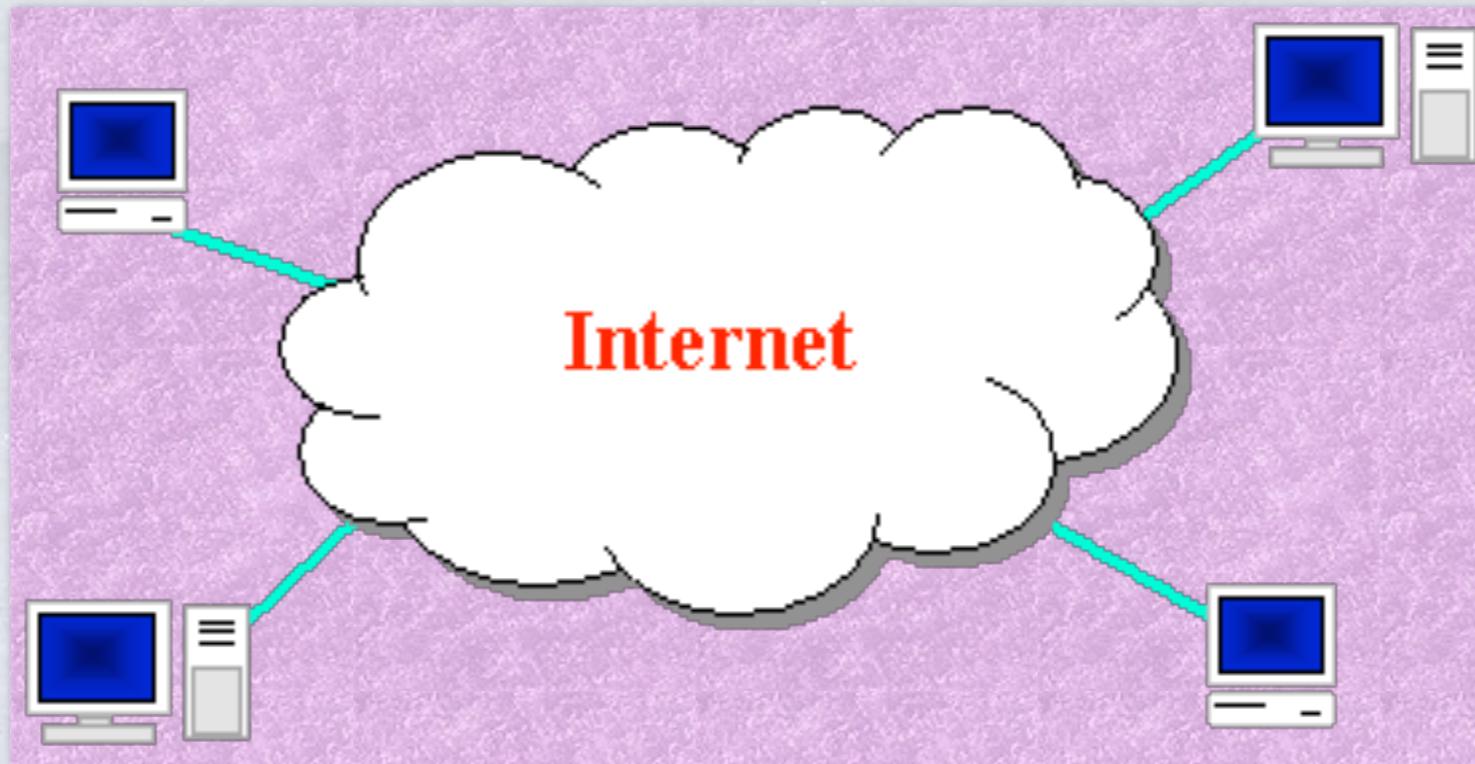
Outline

❖ **The Internet**

❖ **The World Wide Web (WWW)**



The Internet



❖ Wikipedia: <http://en.wikipedia.org/wiki/Internet>

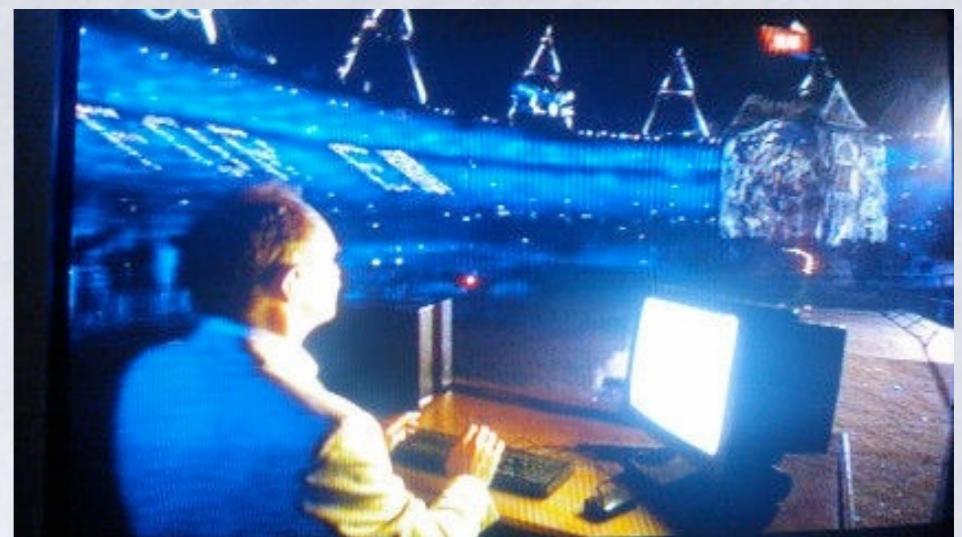
- * a connection of computer networks using the Internet Protocol Suite (TCP/IP)

❖ What's the difference between the Internet and the World Wide Web (WWW)?

- * WWW = HTML* + HTTP(S)
- * including CSS, JavaScript, and other browser enabled content

Brief History

- ❖ WWW created in 1989-91 by Tim Berners-Lee
 - * This is for Everyone
- ❖ popular web browsers released:
 - * Netscape 1994, IE 1995
- ❖ Amazon.com opens in 1995
- ❖ Google January 1996
- ❖ 2000-2002 .COM
- ❖ 2004 web2.0
- ❖ 2016 web3.0



People and Organizations

Internet Engineering Task Force (IETF)

- * internet protocol standards

Internet Corporation for Assigned Names and Numbers (ICANN)

- * decides top-level domain names

World Wide Web Consortium (W3C)

- * Web standards



Internet Protocol (IP)

- ❖ Simple protocol for data exchange between computers
- ❖ The IP is the underlying system of communication for all data (packets) sent across the internet.
 - * Subnetting
 - * Classless Inter Domain Routing (CIDR)
 - * Network Address Translation (NAT)
- ❖ IPv4 vs IPv6

Transmission Control Protocol (TCP)

- ❖ Adds multiple, guaranteed message delivery on top of IP
- ❖ multiplexing: multiple programs using the same IP address
 - * port: a number given to each program or service
 - * 80: Web browser (443 for secure browsing)
 - * 25: email
 - * 22: ssh
 - * 21: ftp
 - * more common ports...
- ❖ some programs (QQ, games, streaming media programs) use simpler UDP protocol instead of TCP
- ❖ find out ports used:
 - * in a terminal, using netstat (Windows) command
 - * using CurrPorts

The Domain Name System

- ❖ The domain name system is usually used to translate a host name into an IP address .
- ❖ Domain names comprise a hierarchy so that names are unique, yet easy to remember.

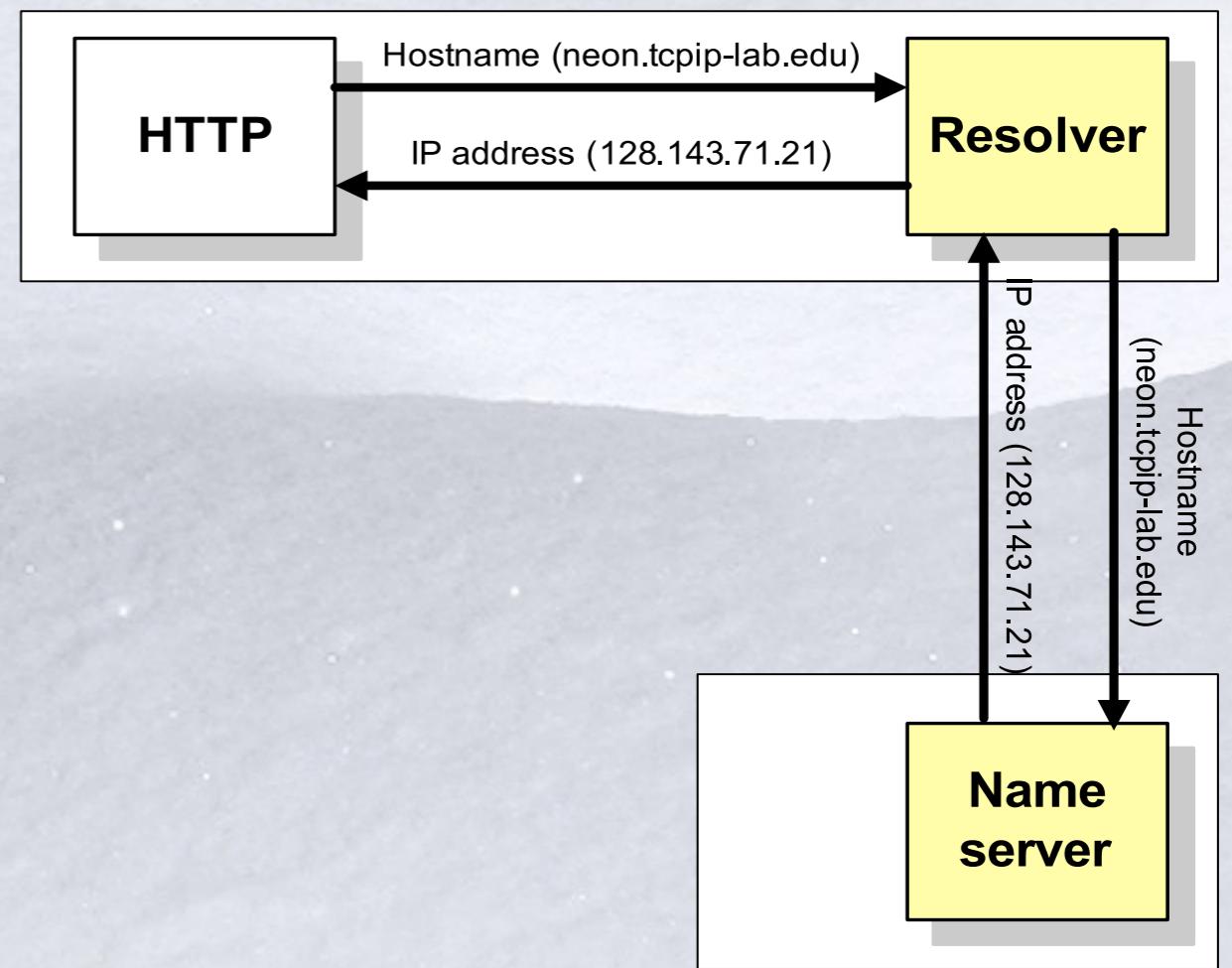
Before there was DNS

.... there was the HOSTS.TXT file

- ✿ Before DNS (until 1985), the name-to-IP address was done by downloading a single file (`hosts.txt`) from a central server with FTP.
 - * Names in `hosts.txt` are not structured.
 - * The `hosts.txt` file still works on most operating systems. It can be used to define local names.
 - * Windows: `C:\Windows\system32\drivers\etc\hosts`
 - * Mac: `/private/etc/hosts`
 - * Linux: `/etc/hosts`

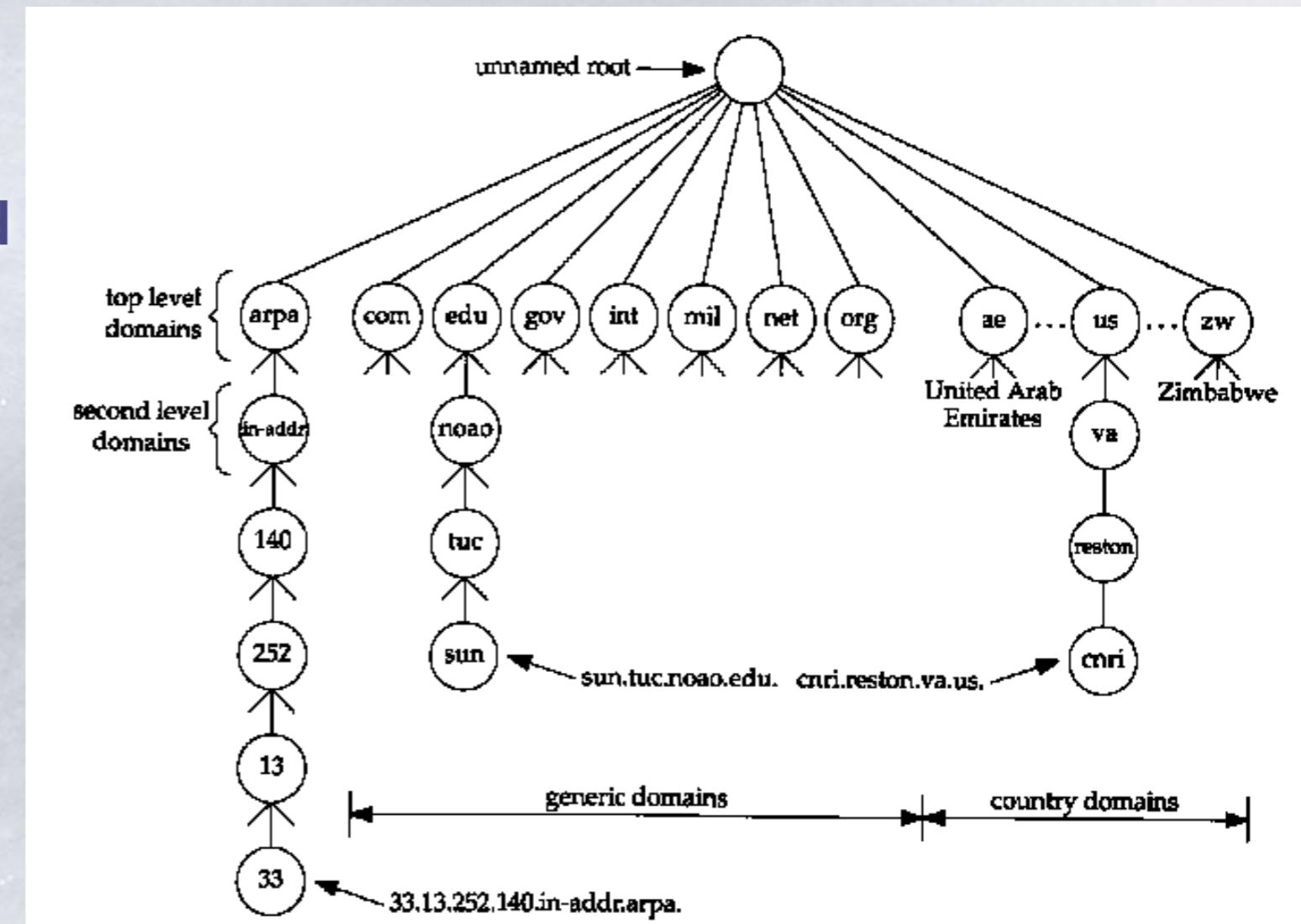
Resolver and Name Server

- ❖ An application program on a host accesses the domain system through a DNS client, called the resolver
- ❖ Resolver contacts DNS server, called name server
- ❖ DNS server returns IP address to resolver which passes the IP address to application
- ❖ Reverse lookups are also possible, i.e., find the hostname given an IP address



DNS Name Hierarchy

- ❖ DNS hierarchy can be represented by a tree
- ❖ Root and top-level domains are administered by an Internet central name registration authority (ICANN)
- ❖ Below top-level domain, administration of name space is delegated to organizations
- ❖ Each organization can delegate further



Domain names

- ❖ Hosts and DNS domains are named based on their position in the domain tree
- ❖ Every node in the DNS domain tree can be identified by a unique Fully Qualified Domain Name (FQDN). The FQDN gives the position in the DNS tree.

software.nju.edu.cn or software.nju.edu.cn.

- ❖ A FQDN consists of labels (“software”, “nju”, “edu”, “cn”) separated by a period (“.”)
 - * There can be a period (“.”) at the end.
 - * Each label can be up to 63 characters long
 - * FQDN contains characters, numerals, and dash character (“-”)
 - * FQDNs are not case-sensitive

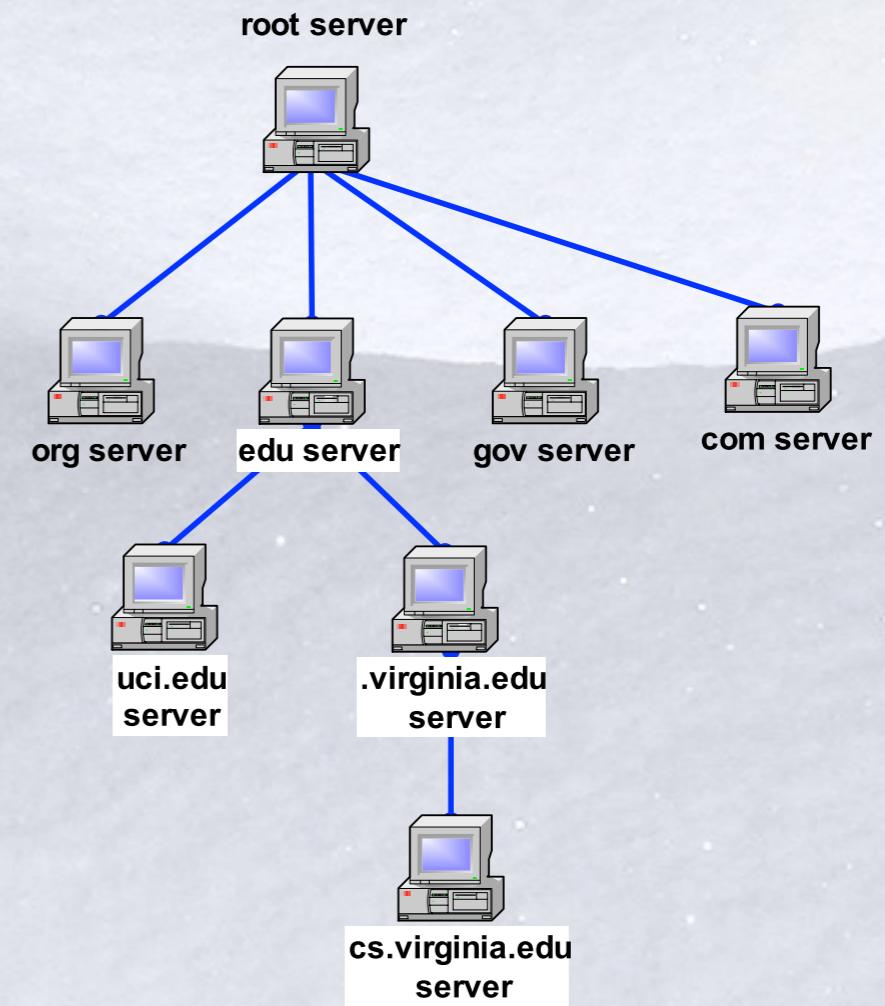
Hierarchy of name servers

- ❖ The resolution of the hierarchical name space is done by a hierarchy of name servers

- ❖ Each server is responsible (authoritative) for a contiguous portion of the DNS namespace, called a zone.

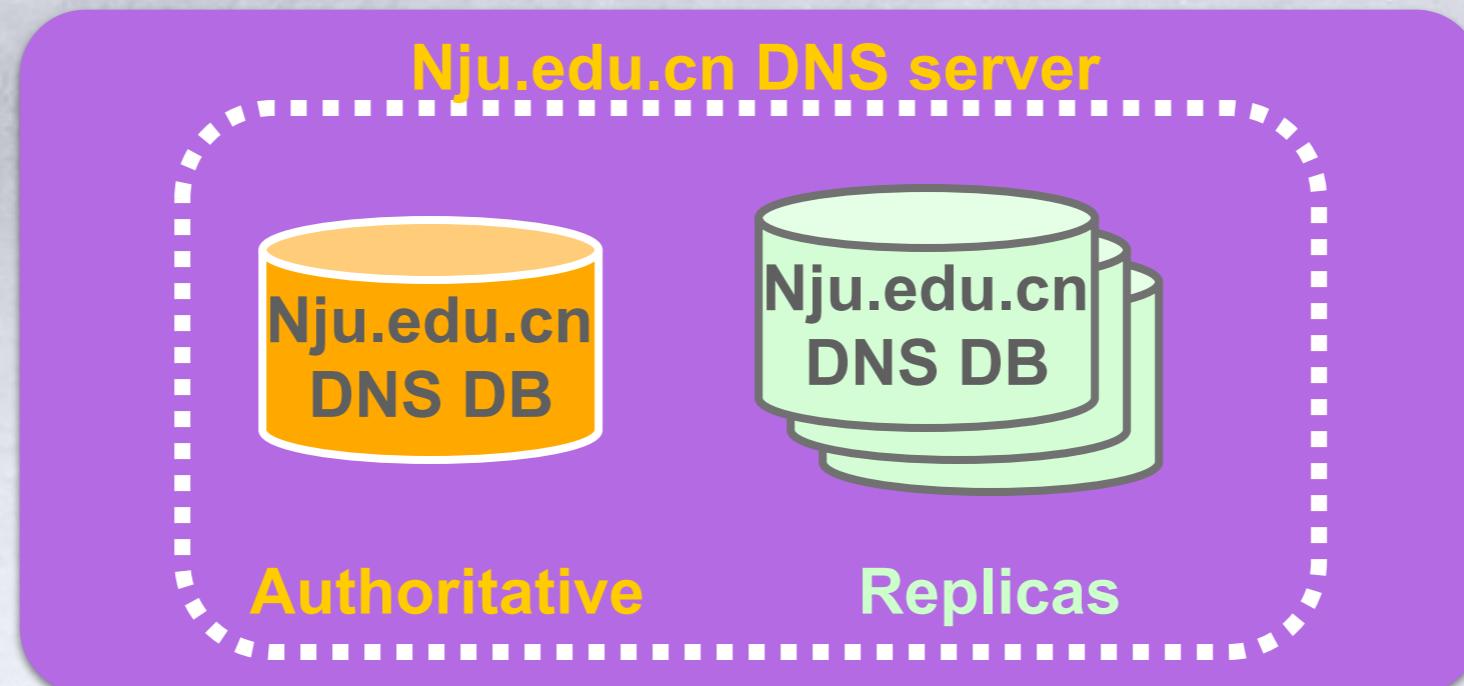
- ❖ Zone is a part of the subtree

- ❖ DNS server answers queries about hosts in its zone



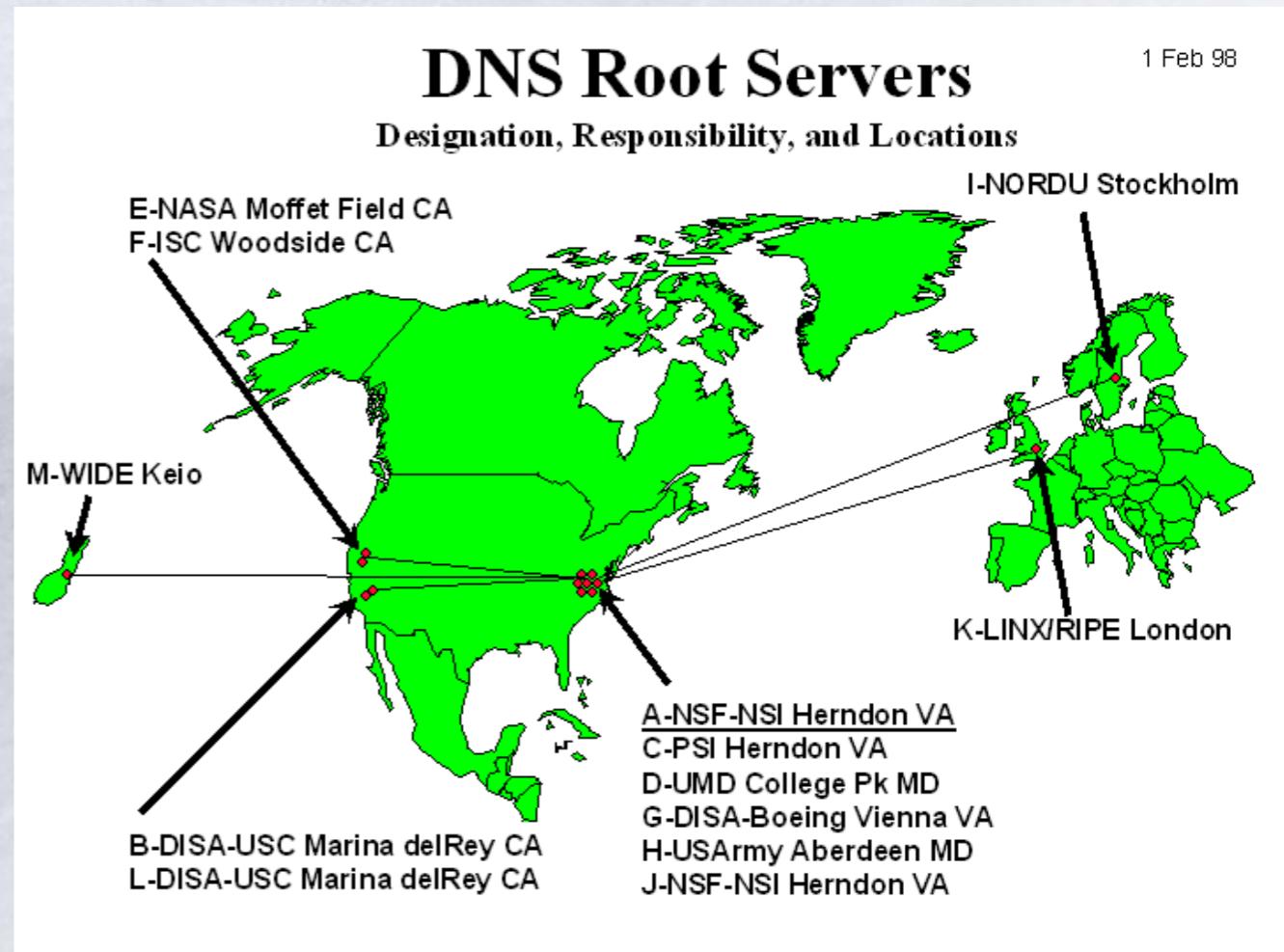
Primary and secondary name servers

- ❖ For each zone, there must be a primary name server and a secondary name server
 - * The primary server (master server) maintains a zone file which has information about the zone. Updates are made to the primary server
 - * The secondary server copies data stored at the primary server.
- ❖ Adding a host:
 - * When a new host is added (“gold.cs.xxx.edu.cn”) to a zone, the administrator adds the IP information on the host (IP address and name) to a configuration file on the primary server



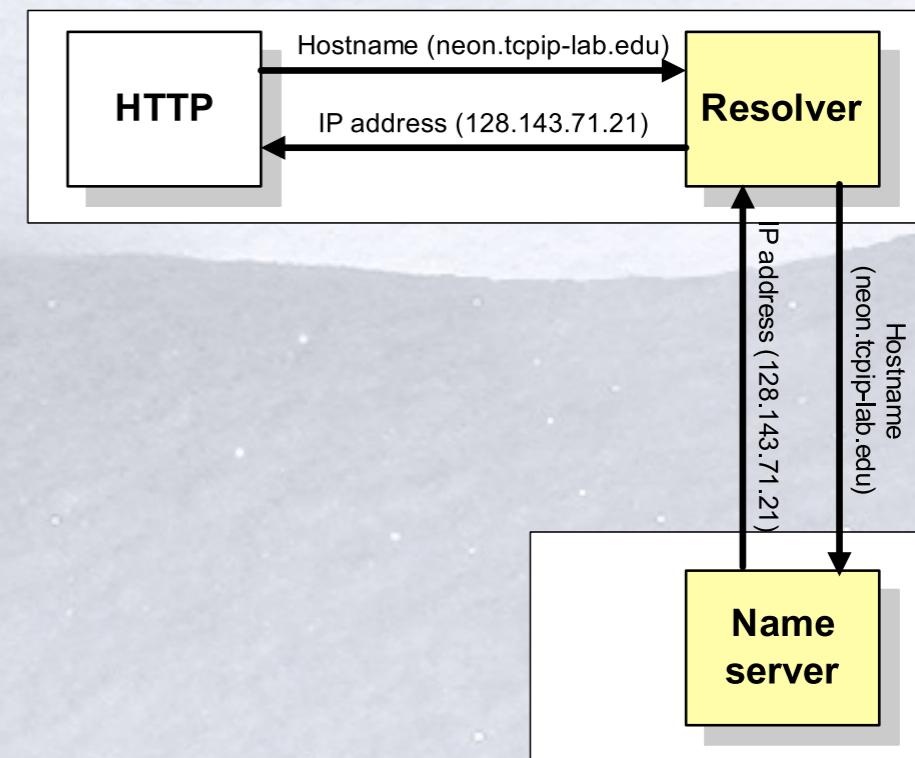
Root name servers

- ❖ The root name servers know how to find the authoritative name servers for all top-level zones.
- ❖ There are only 13 root name servers
- ❖ Root servers are critical for the proper functioning of name resolution



Domain name resolution

- ❖ User program issues a request for the IP address of a hostname
- ❖ Local resolver formulates a DNS query to the name server of the host
- ❖ Name server checks if it is authorized to answer the query.
 - * If yes, it responds.
 - * Otherwise, it will query other name servers, starting at the root tree
- ❖ When the name server has the answer it sends it to the resolver.



Recursive and Iterative Queries

- ❖ There are two types of queries:
 - * Recursive queries
 - * Iterative (non-recursive) queries
- ❖ The type of query is determined by a bit in the DNS query
- ❖ Recursive query: When the name server of a host cannot resolve a query, the server issues a query to resolve the query
- ❖ Iterative queries: When the name server of a host cannot resolve a query, it sends a referral to another server to the resolver

Caching

- ❖ To reduce DNS traffic, name servers caches information on domain name/IP address mappings
- ❖ When an entry for a query is in the cache, the server does not contact other servers

Resource Records

- ❖ The database records of the distributed data base are called resource records (RR)
- ❖ Resource records are stored in configuration files (zone files) at name servers.
- ❖ DNS databases contain more than just hostname-to-address records:
 - * Name server records NS
 - * Hostname aliases CNAME
 - * Mail Exchangers MX
 - * Host Information HINFO

Resource Identifiers

URI (Uniform Resource Identifier)

- * permits resources to reside anywhere on the Internet

URL (Uniform Resource Locator)

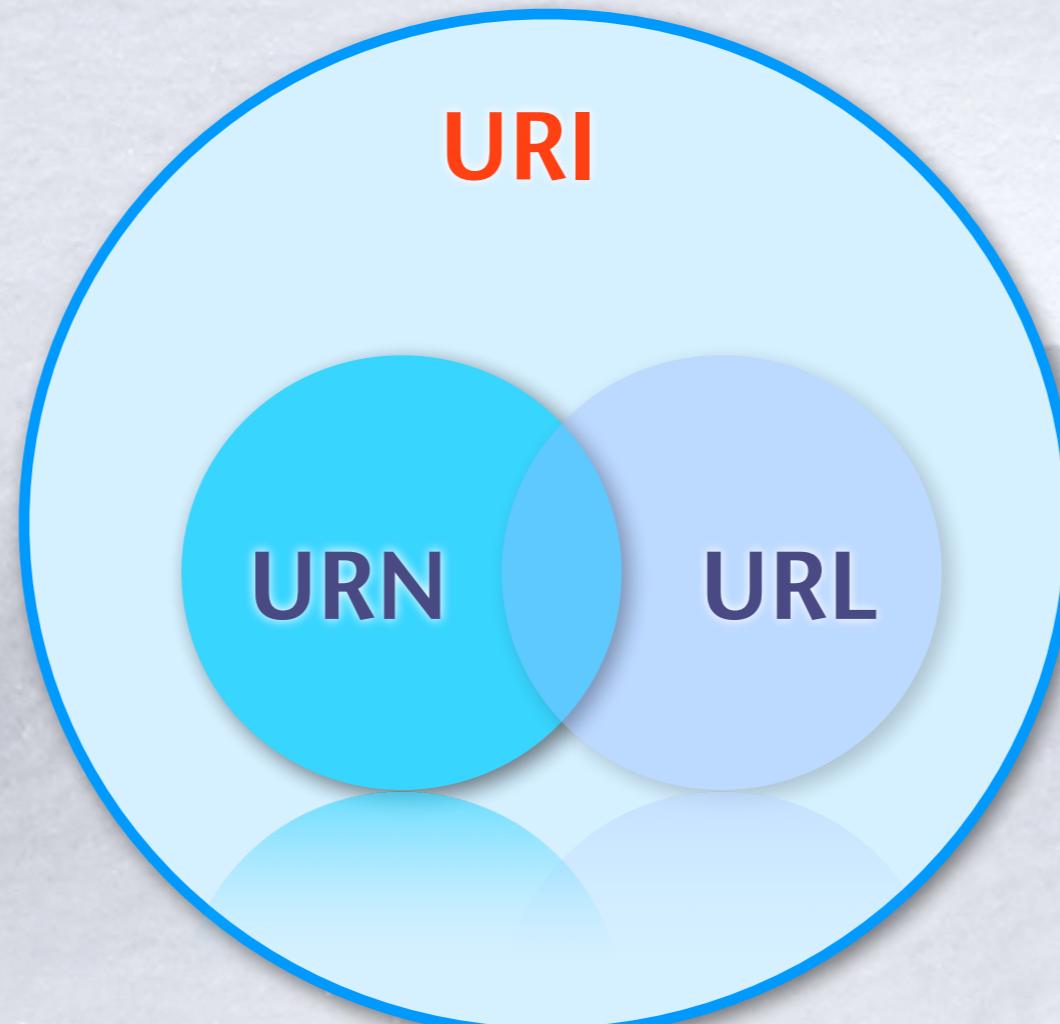
- * shows the location of a copy of a resource

URN (Uniform Resource Name)

- * is a unique name for a resource

URI

- * URI is a superset of both URL and URN
- * URIs defined in RFC 2396.
- *



URN

❄ URN (uniform resource name)

- * these location-independent URNs allow resources to move from place to place

❄ URN need a supporting infrastructure to resolve resource locations, that's why it is not yet widely adopted.

❄ Example URN

`urn:ietf:rfc:2141`

Hypertext

Text with links to other text

- * Click on links takes you somewhere else

Web adapted the idea, link specification:

- * Uniform Resource Locators (URL) - Provided names for web content

```
<a href="https://en.wikipedia.org/wiki/URL">URL</a>
```

Parts of an URL

`http://host.company.com:80/a/b/c.html?user=Alice&year=2018#p2`

- ❖ **Scheme (http:): identifies protocol used to fetch the content.**
- ❖ **Host name (//host.company.com): name of a machine to connect to.**
- ❖ **Server's port number (80): allows multiple servers to run on the same machine.**
- ❖ **Hierarchical portion (/a/b/c.html): used by server to find content.**
- ❖ **Query parameters (?user=Alice&year=2008): provides additional parameters**
- ❖ **Fragment (#p2): Have browser scroll page to fragment (html: p2 is anchor tag)**

URL: schemes (e.g. http)

- ❖ http: is the most common scheme; it means use the HTTP protocol
- ❖ https: is similar to http: except that it uses SSL encryption
- ❖ file: means read a file from the local disk
- ❖ mailto: means open an email program composing a message
- ❖ There are several other schemes, such as ftp:, but they aren't used much anymore.

URL: Hierarchical portion (/a/b/c.html)

❖ Passed to the web server for interpretation. Early web servers:

- * Path name for a static HTML file.
- * Path name of a program that will generate the HTML content (e.g., foo.php).

❖ Web server programmed with routing information

- * Map hierarchical position to function to be performed and possibly the function's parameters

❖ Application Programming Interface (API) design,
Example:

- * /user/create
- * /user/list
- * /user/0x23490
- * /user/0x23433
- * /user/delete/0x23433

Query Parameters (e.g. ?user=Alice&year=2008)

- ❖ Traditionally has been to provide parameters to operation:

`http://www.company.com/showOrder.php?order=4621047`

- ❖ For modern apps has implications of when the browser switches pages

Links

- ❄ Browser maintains a notion of current location (i.e. URL)
- ❄ Links: content in a page which, when clicked on, causes the browser to go to URL
- ❄ Links are implemented with the <a> tag:
`2009 News`

Different types of links

- ❖ Full URL: <a href="<http://www.xyz.com/news/2019.html>">2019 News
- ❖ Absolute URL:
 - * same as <http://www.xyz.com/stock/quote.html>
- ❖ Relative URL (intra-site links):
 - * same as <http://www.xyz.com/news/2018/March.html>
- ❖ Define an anchor point (a position that can be referenced with # notation):
 - * Go to a different place in the same page:

Uses of URLs

Loading a page: type the URL into your browser

- * Load a image:

```

```

- * Load a stylesheet:

```
<link rel="stylesheet" type="text/css" href="...">
```

- * Embedded a page:

```
<iframe src="http://www.google.com">
```

URL Encoding

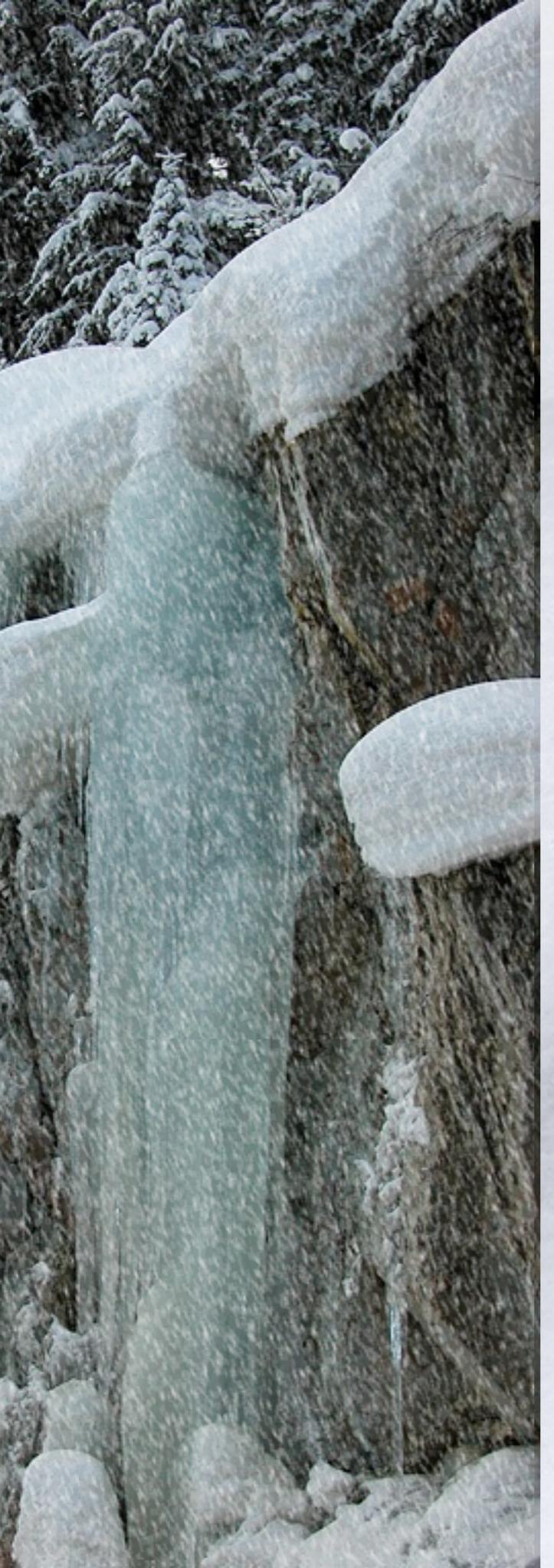
- ❖ What if you want to include a punctuation character in a query value?

`http://www.stats.com/companyInfo?name=C&H Sugar`

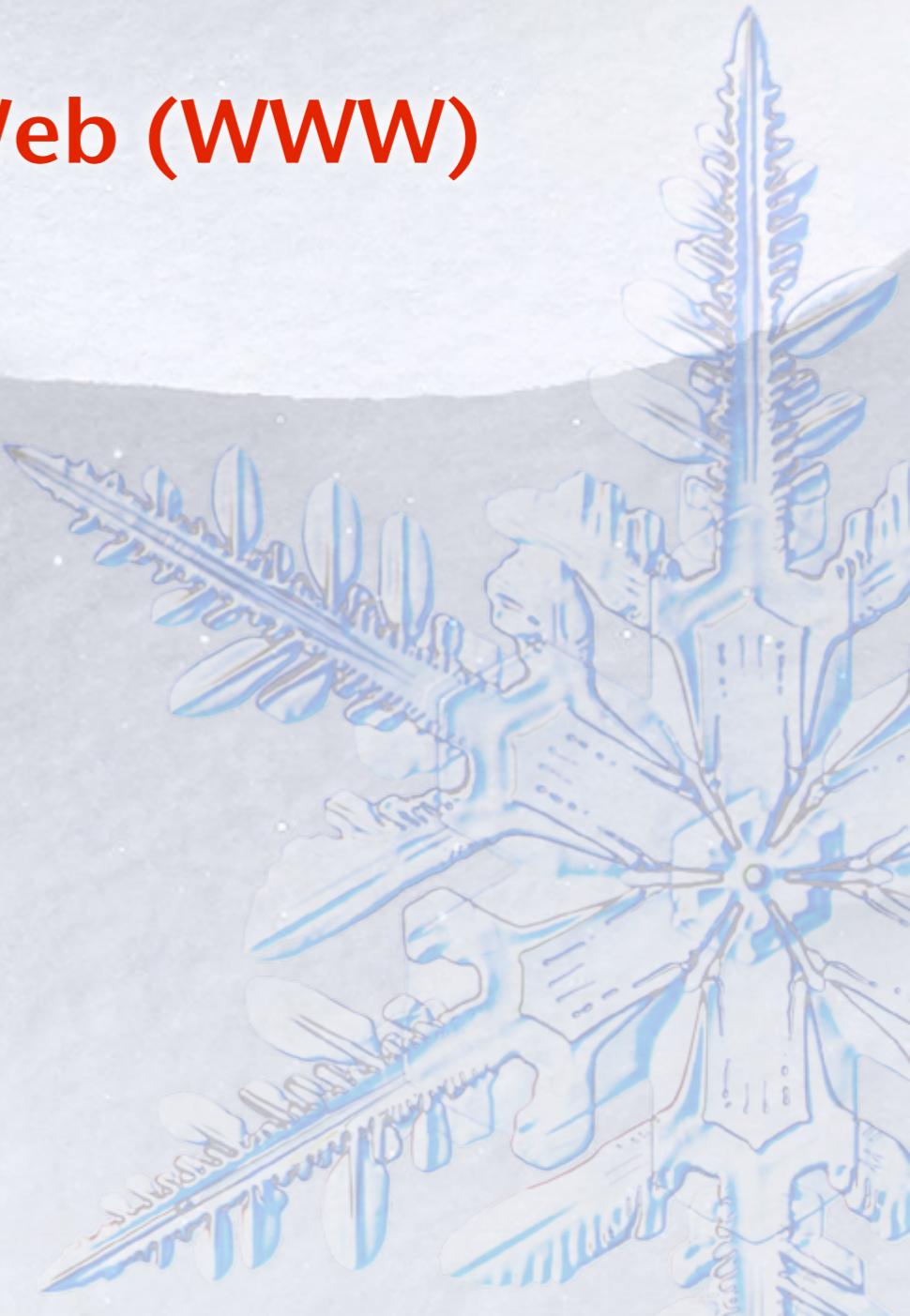
- ❖ Any character in a URL other than A-Z, a-z, 0-9, or any of -_.~ must be represented as %xx, where xx is the hexadecimal value of the character:

`http://www.stats.com/companyInfo?
name=C%26H%20Sugar`

- ❖ Escaping is a commonly used technique and also a source of errors



Outline

- ❖ The Internet
 - ❖ The World Wide Web (WWW)
- 

Web servers and browsers

❖ Web server: software that listens for Web page requests

- * Apache
- * Microsoft Internet Information Server (IIS) (part of Windows)

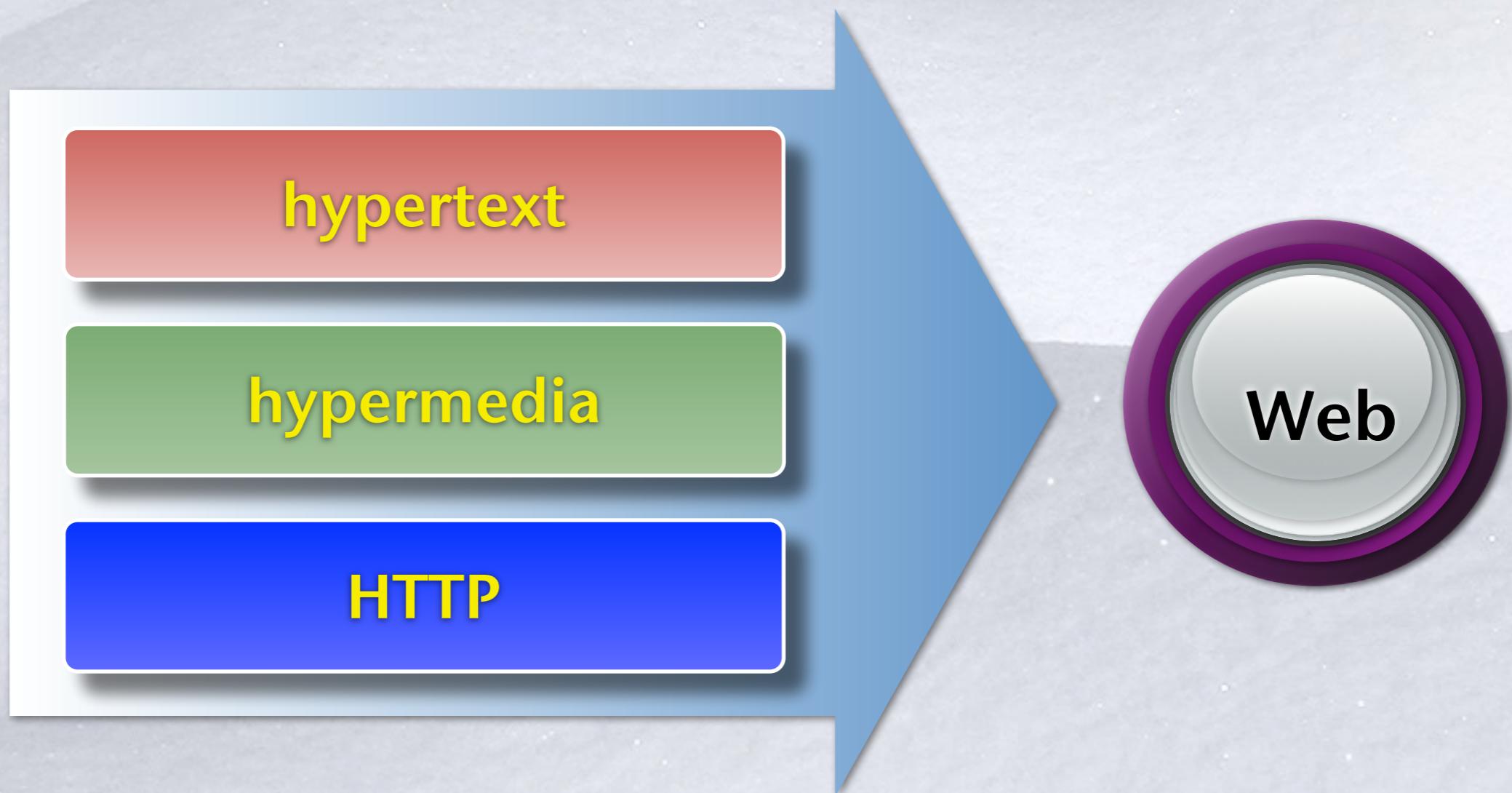


❖ Web browser: fetches/displays documents from Web servers

- * Microsoft Internet Explorer (IE)
- * Mozilla Firefox
- * Apple Safari
- * Google Chrome
- * Opera



Web的表现形式



互联网的影响

❄ 改变生活方式

- * 说走就走的旅行

❄ 改变交流方式

- * 短信 vs 微信

❄ 改变购物方式

- * 随时随地、在线购物、移动支付

❄ 改变娱乐方式

- * 团购、在线选座 / 订位、电子门票、手机游戏

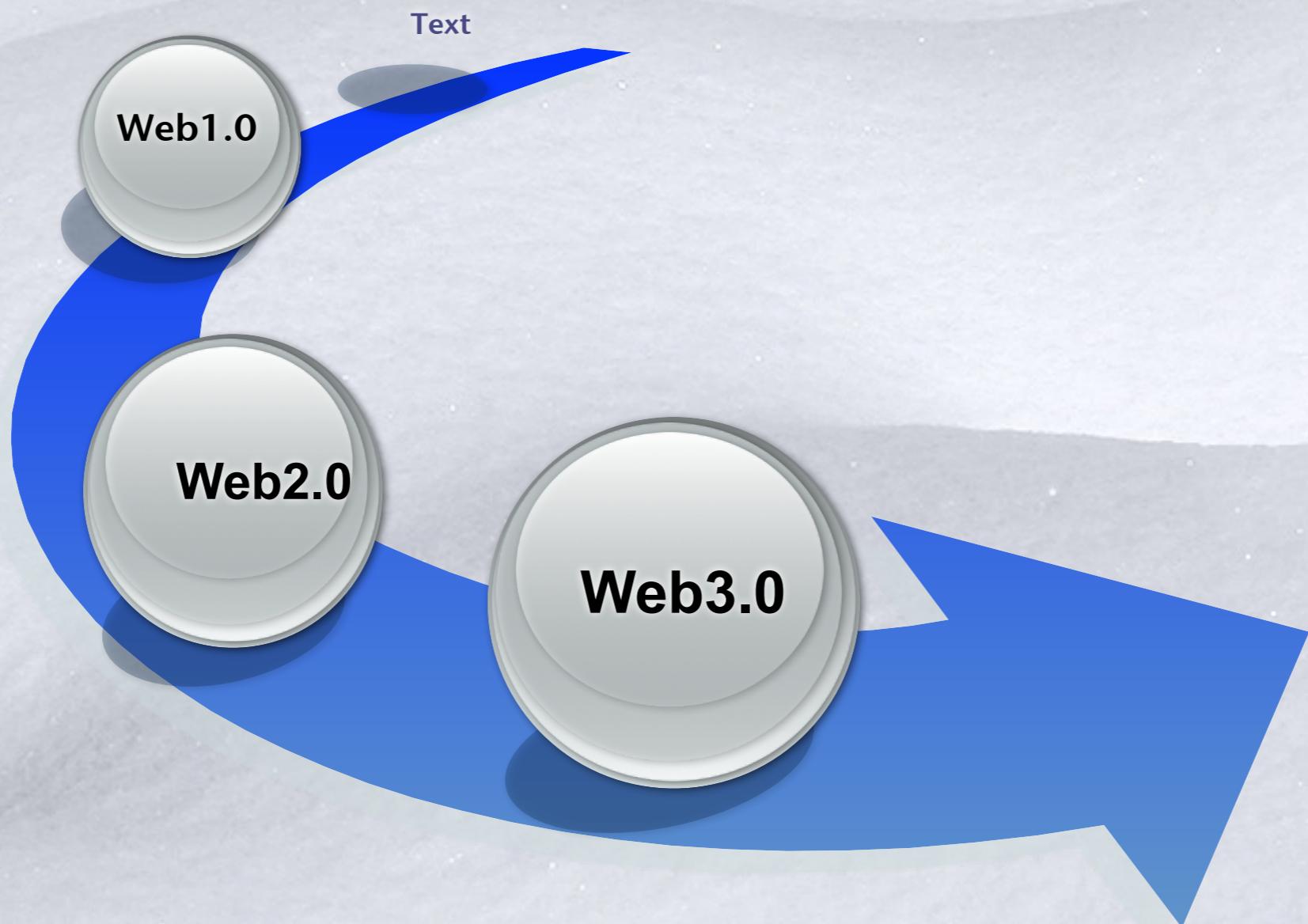
❄ 改变教育方式

- * mooc

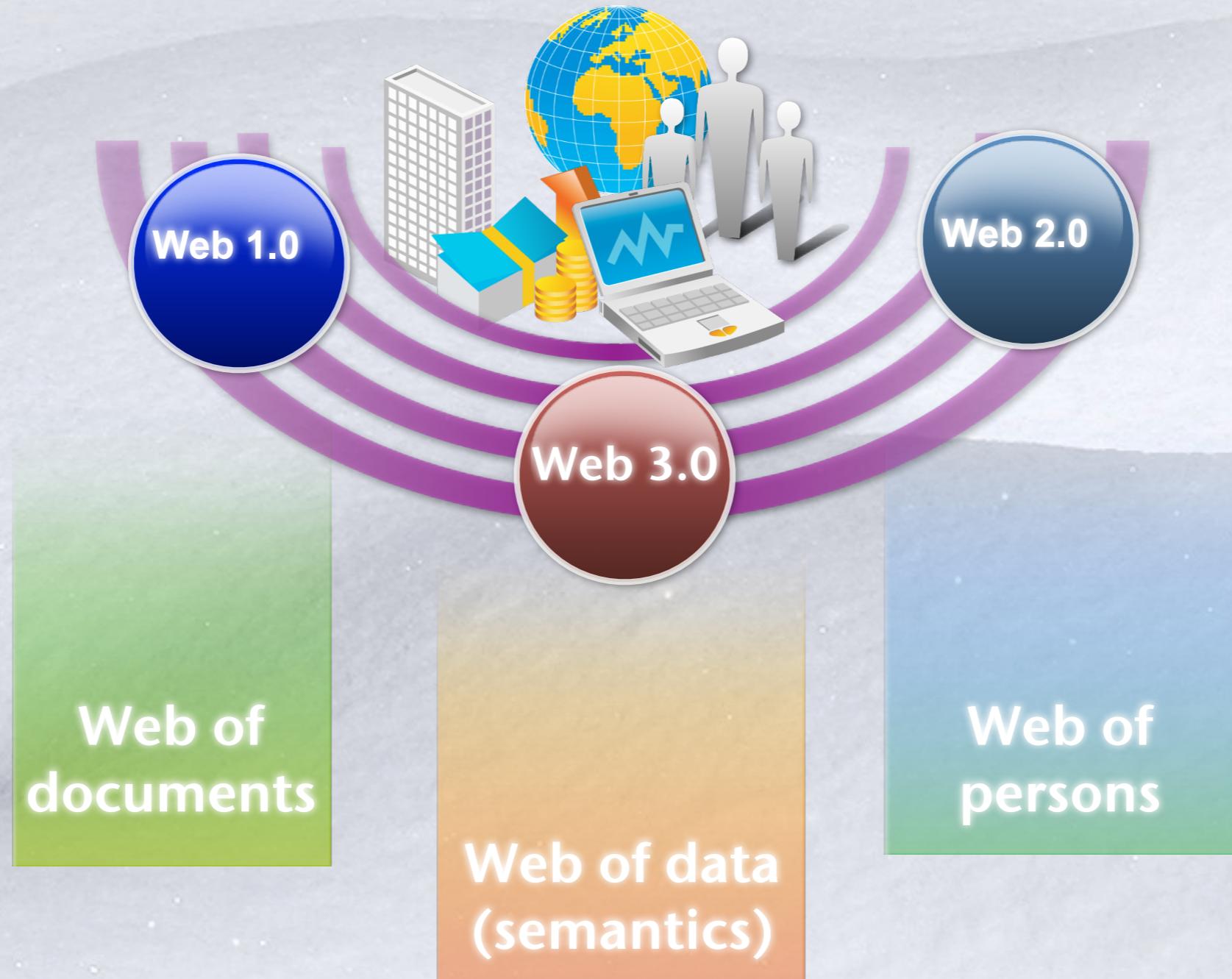
❄ 改变医疗方式

- * 远程医疗、网售药品

Web发展历程



Web1.0 Web2.0 Web 3.0



Web 2.0 特点

- * Web2.0是一个架构在知识上的环境，人与人之间交互而产生出的内容，经由在服务导向的架构中的程序，在这个环境中被发布、管理和使用。

IBM的社区网络分析师Dario de Judicibus



Web 2.0 典型应用

- ❄ Facebook (meta)
- ❄ Wiki百科
- ❄ 豆瓣
- ❄ 土豆网
- ❄ 微博

Web2.0的局限性

- ❄ 过度饱和
- ❄ 概念偏差
- ❄ 时间
- ❄ 互动模式
- ❄ 开放

“人们不停地质问Web 3.0到底是什么。我认为当可缩放矢量图形在Web 2.0的基础上大面积使用——所有东西都起波纹、被折叠并且看起来没有棱角——以及一整张语义网涵盖著大量的数据，你就可以访问这难以置信的数据资源。”

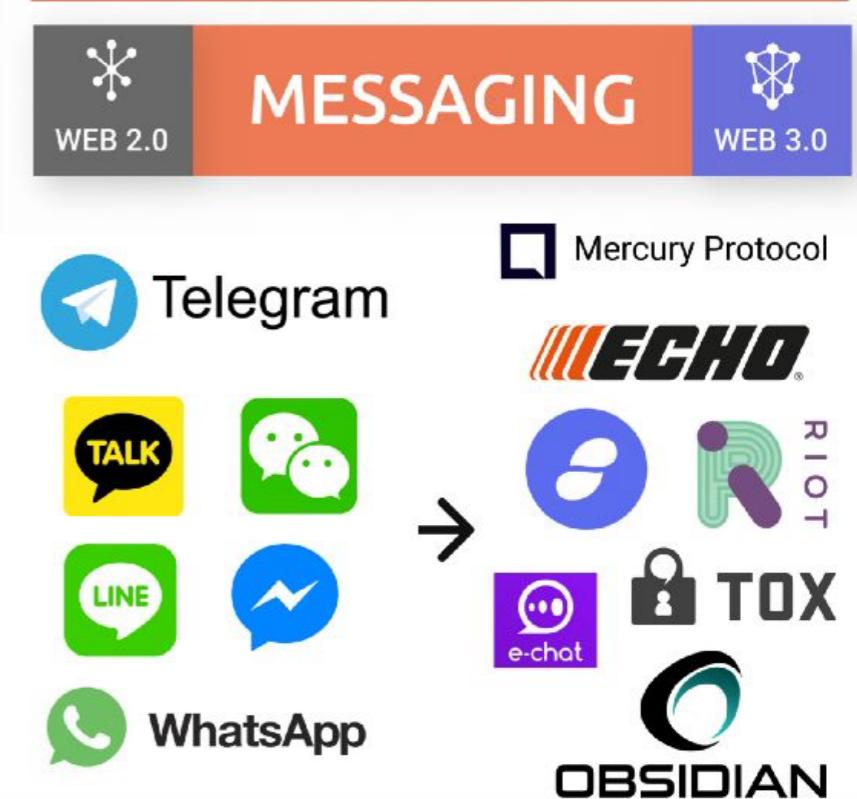
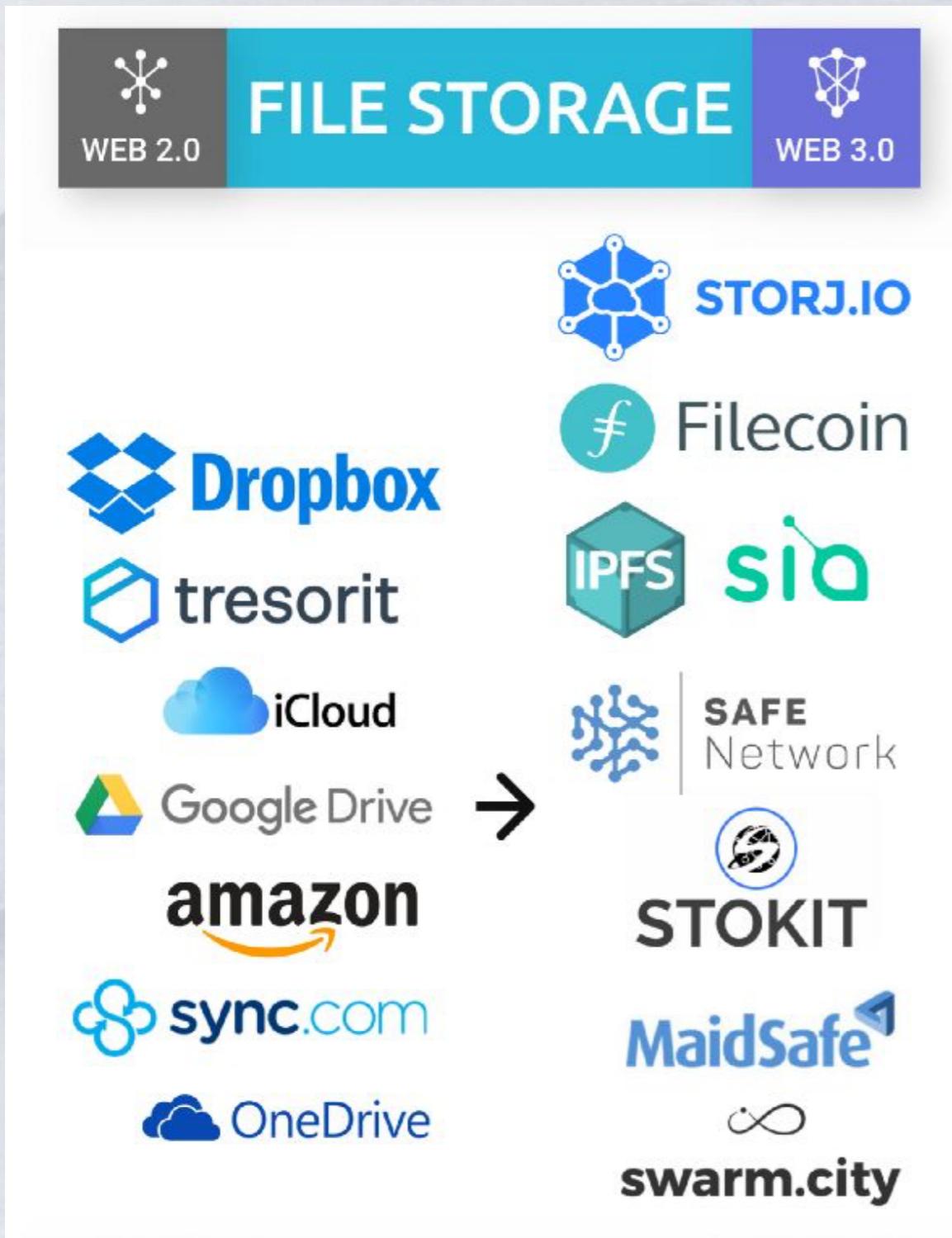
——Tim Berners-Lee

Victoria Shannon. . A 'more revolutionary' Web. : International Herald Tribune., 2006-06-26

Web3.0特征

- ❄ 语义化
- ❄ 3D
- ❄ 人工智能
- ❄ 去中心化

对比



媒体



Reading materials

- ❄ A Brief History of the Internet <http://www.isoc.org/internet/history/brief.shtml>
- ❄ http://en.wikipedia.org/wiki/Web_2.0
- ❄ <http://oreilly.com/web2/archive/what-is-web-20.html>
- ❄ https://en.wikipedia.org/wiki/Semantic_Web#Web_3.0

Thanks!!!

