The Web

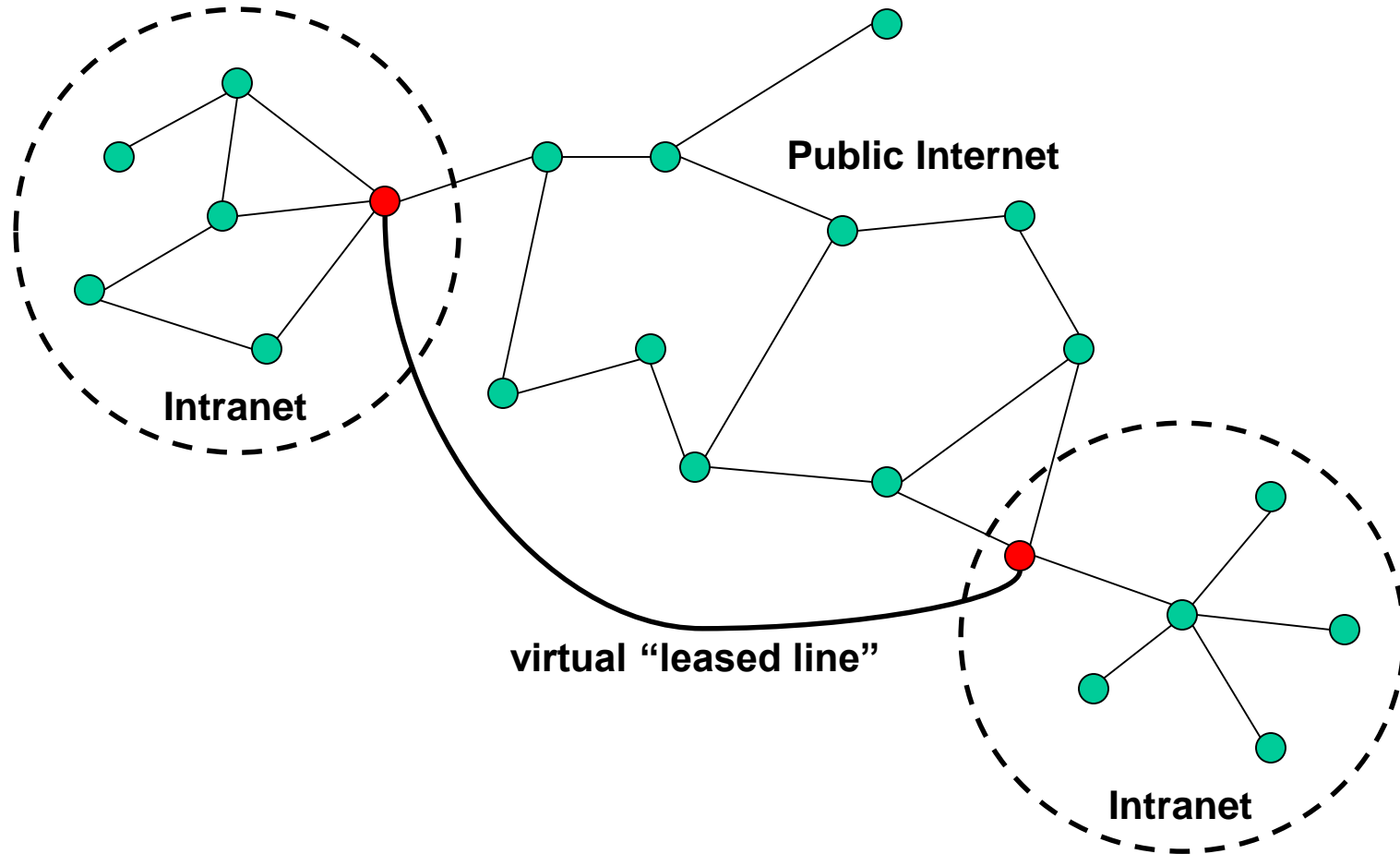## Week 10

LBSC 671

Creating Information Infrastructures

Virtual Private Networks

a secure private network over the public Internet

Public Internet

Intranet

Intranet

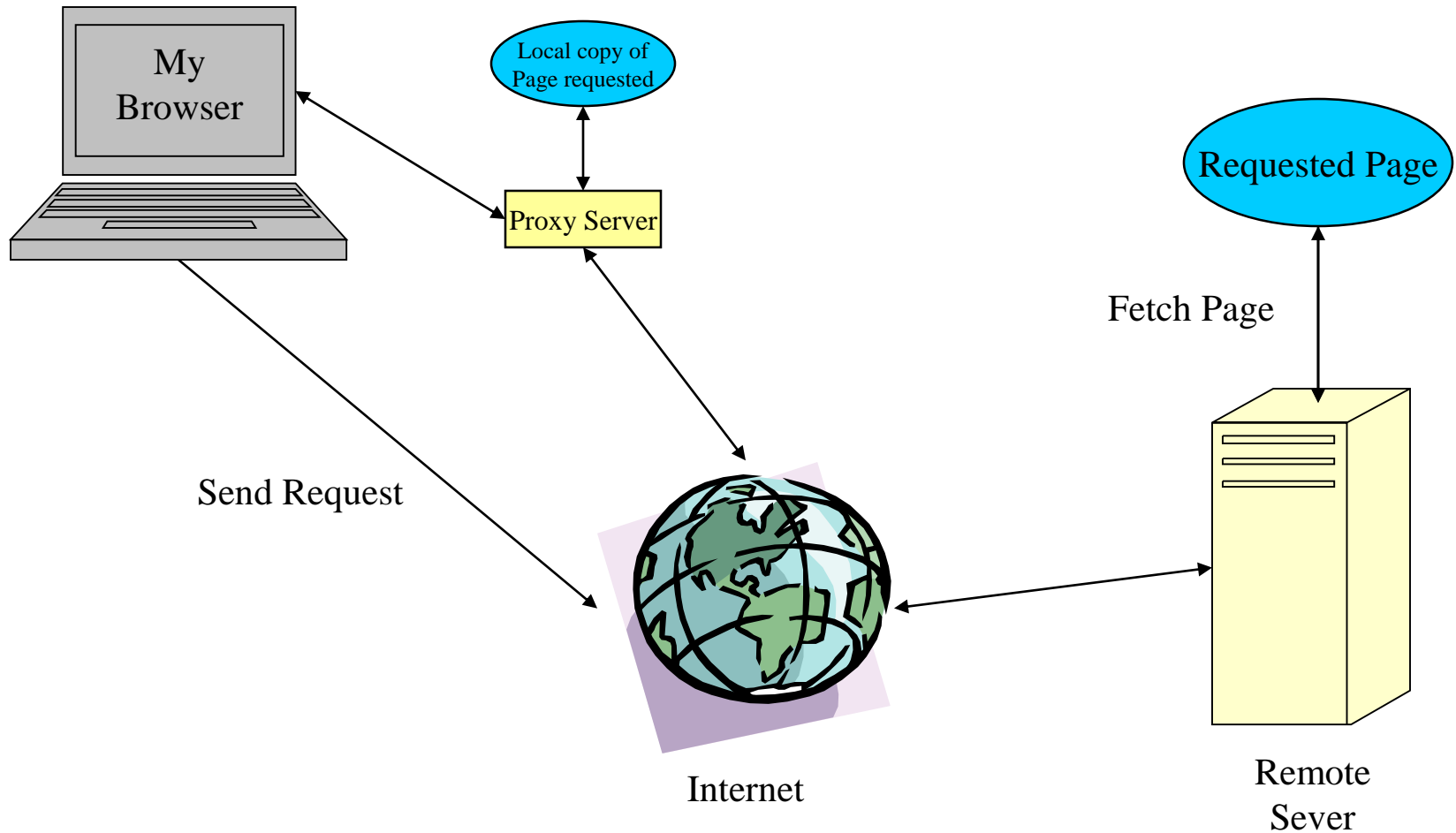**virtual "leased line"**

Tonight

- Think about what the Web "is"

- Learn to create a Web page

- Begin to think about Web design

- Talk conceptually about databases
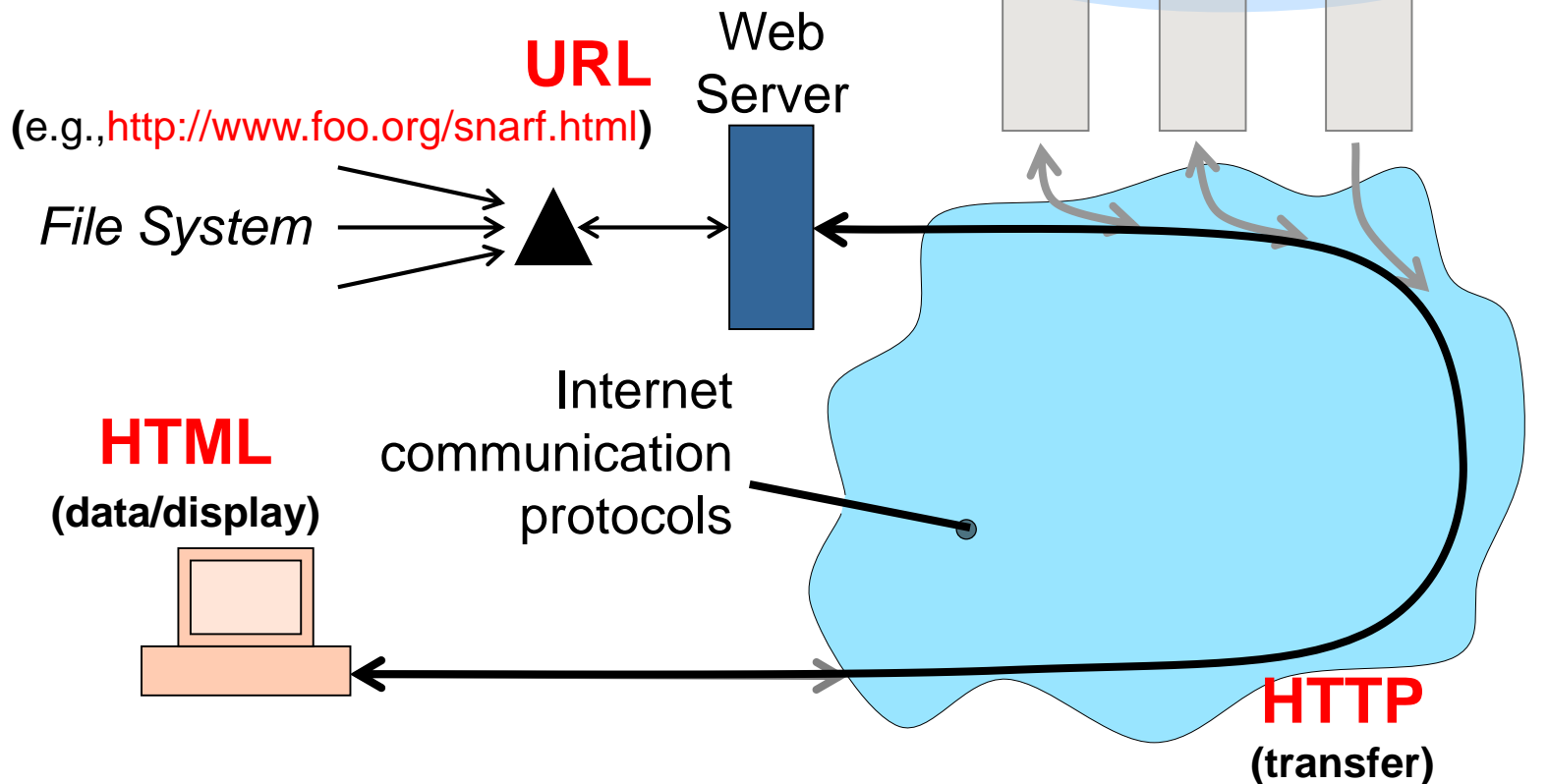
Internet ≠ Web

- Internet: collection of global networks

- Web: way of managing information exchange

- There are many other uses for the Internet
  - File transfer (FTP)
  - Email (SMTP, POP, IMAP)

The World-Wide Web

My
Browser

Local copy of
Page requested

Proxy Server

Requested Page

Fetch Page

Send Request

Internet

Remote
Sever

"The Web"

HTML
HTTP
URL

Email   FTP   RTSP

URL

(e.g.,http://www.foo.org/snarf.html)

File System

Web
Server

HTML

(data/display)

Internet
communication
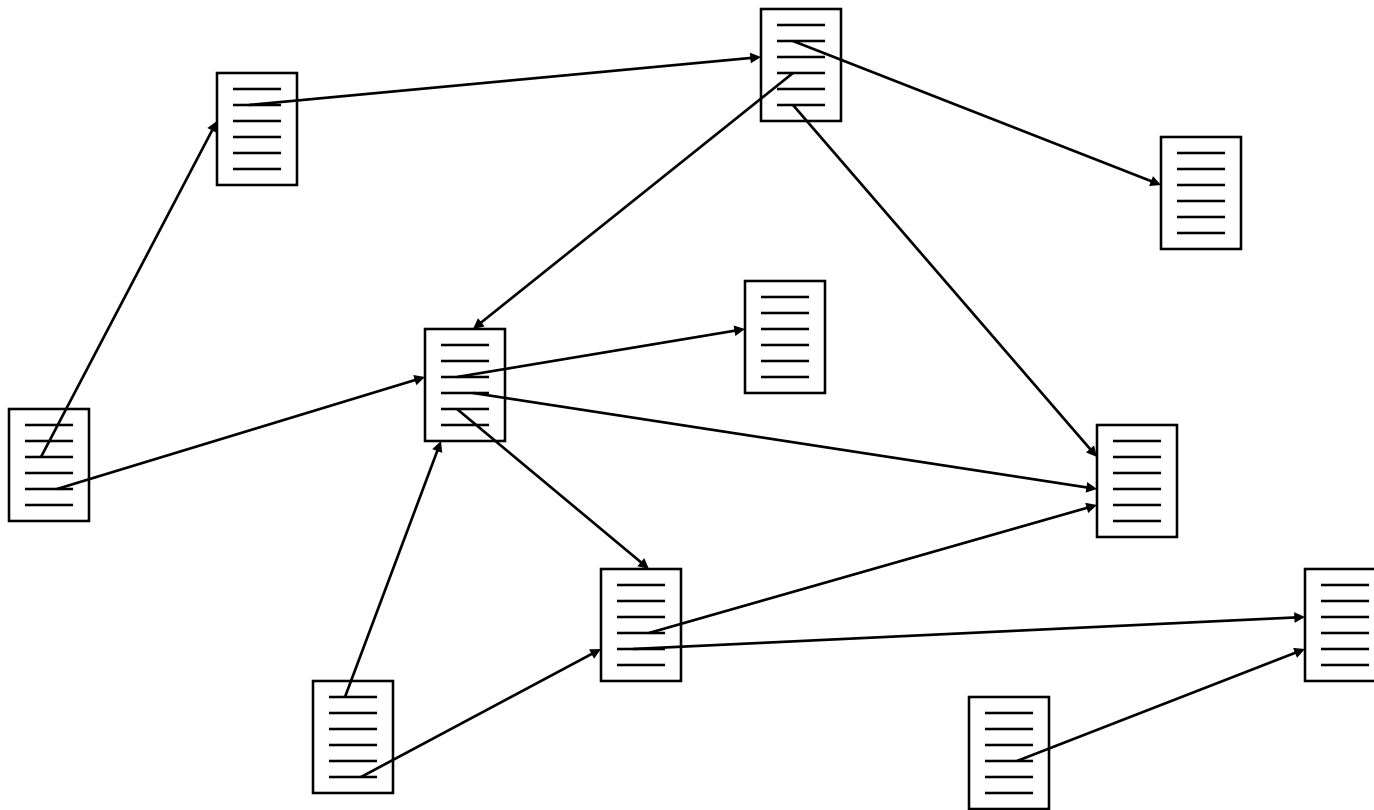protocols

HTTP

(transfer)

What is the Web?

- Protocols
  - HTTP, HTML, or URL?
- Perspective
  - Content or behavior?
- Content
  - Static, dynamic or streaming?
- Access
  - Public, protected, or internal?

Perspectives on the Web

- ## Web "sites"
  - In 2002, OCLC counted any server at port 80
  - Total was 3 million, an undercount
    - Misses many servers at other ports
    - Some servers host unrelated content (e.g., TerpConnect)

- ## Web "pages"
  - In 2012, Google counted any URL it has seen
  - Total was 30 trillion, an overcount
    - Includes dead links, spam, …

- ## Web "use"
  - Google users pose 3 billion queries a day

Crawling the Web

Robots Exclusion Protocol

- Requires voluntary compliance by crawlers

- Exclusion by site
  - Create a robots.txt file at the <u>server's</u> top level
  - Indicate which directories not to crawl

- Exclusion by document (in HTML head)
  - Not implemented by all crawlers

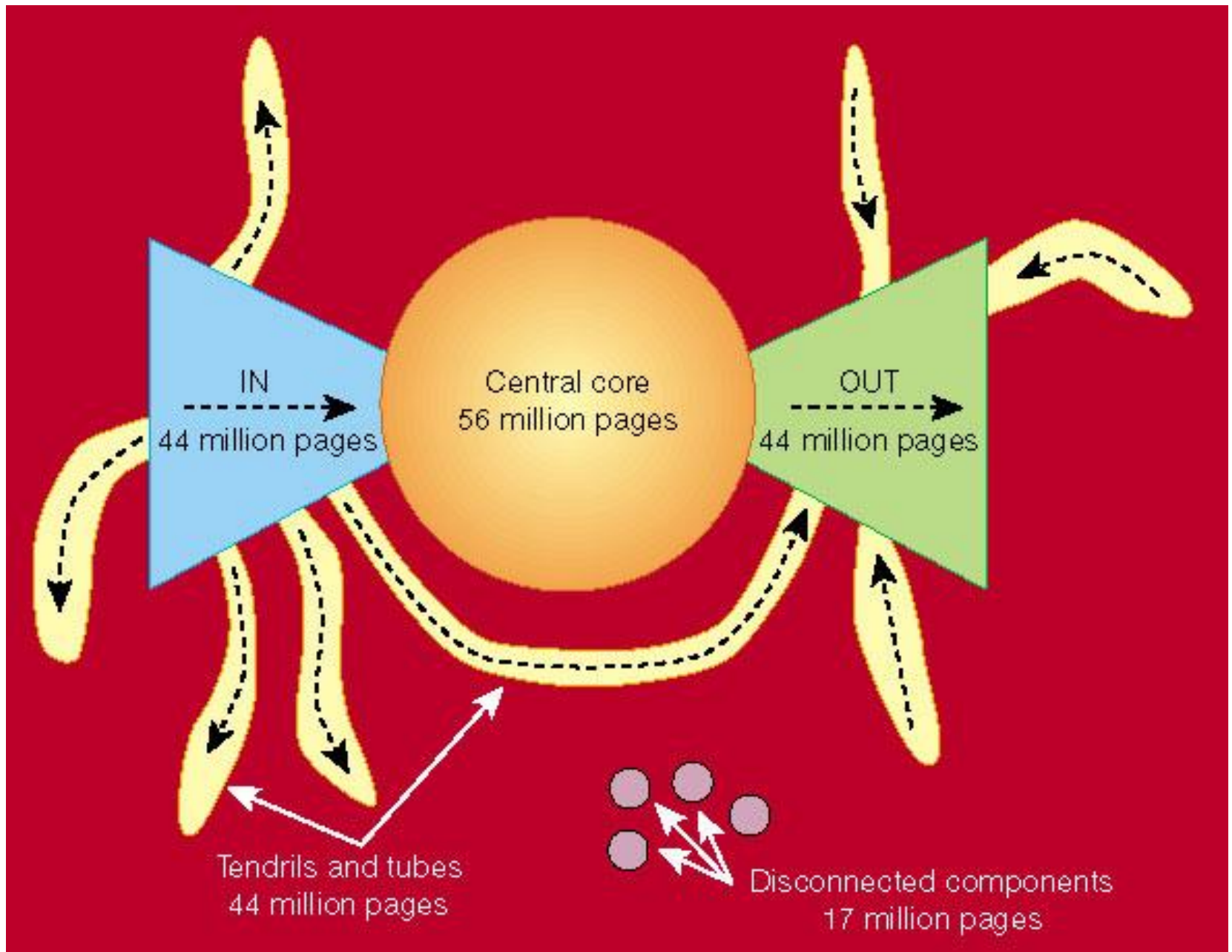   <meta name="robots" content="noindex,nofollow">

Web Crawl Challenges

- Discovering "islands" and "peninsulas"

- Duplicate and near-duplicate content
  - 30-40% of total content

- Link rot
  - Changes at ~1% per week

- Network instability
  - Temporary server interruptions
  - Server and network loads

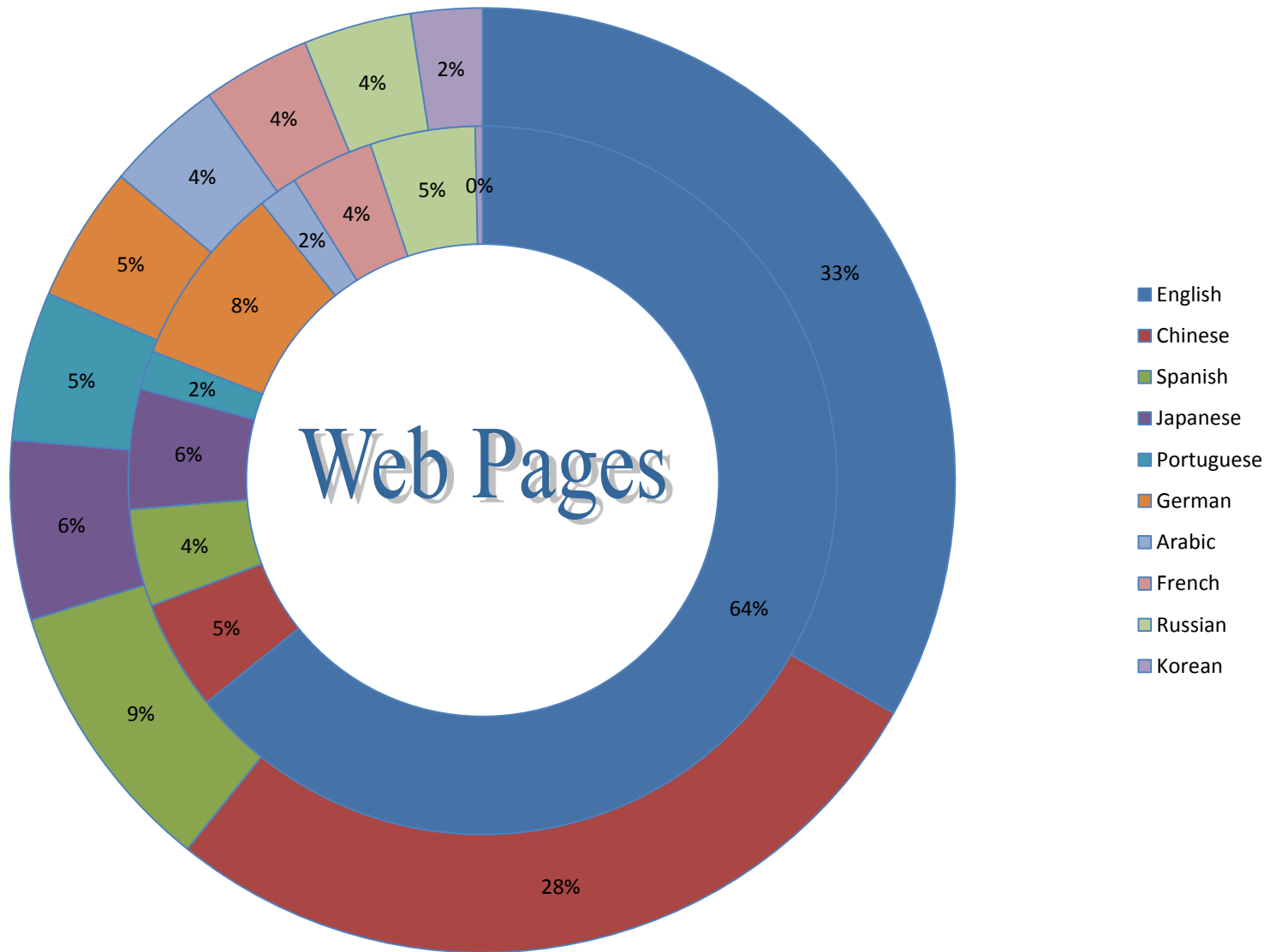- Dynamic content generation

Duplicate Detection

- Structural
  - Identical directory structure (e.g., mirrors, aliases)

- Syntactic
  - Identical bytes
  - Identical markup (HTML, XML, …)

- Semantic
  - Identical content
  - Similar content (e.g., with a different banner ad)
  - Related content (e.g., translated)
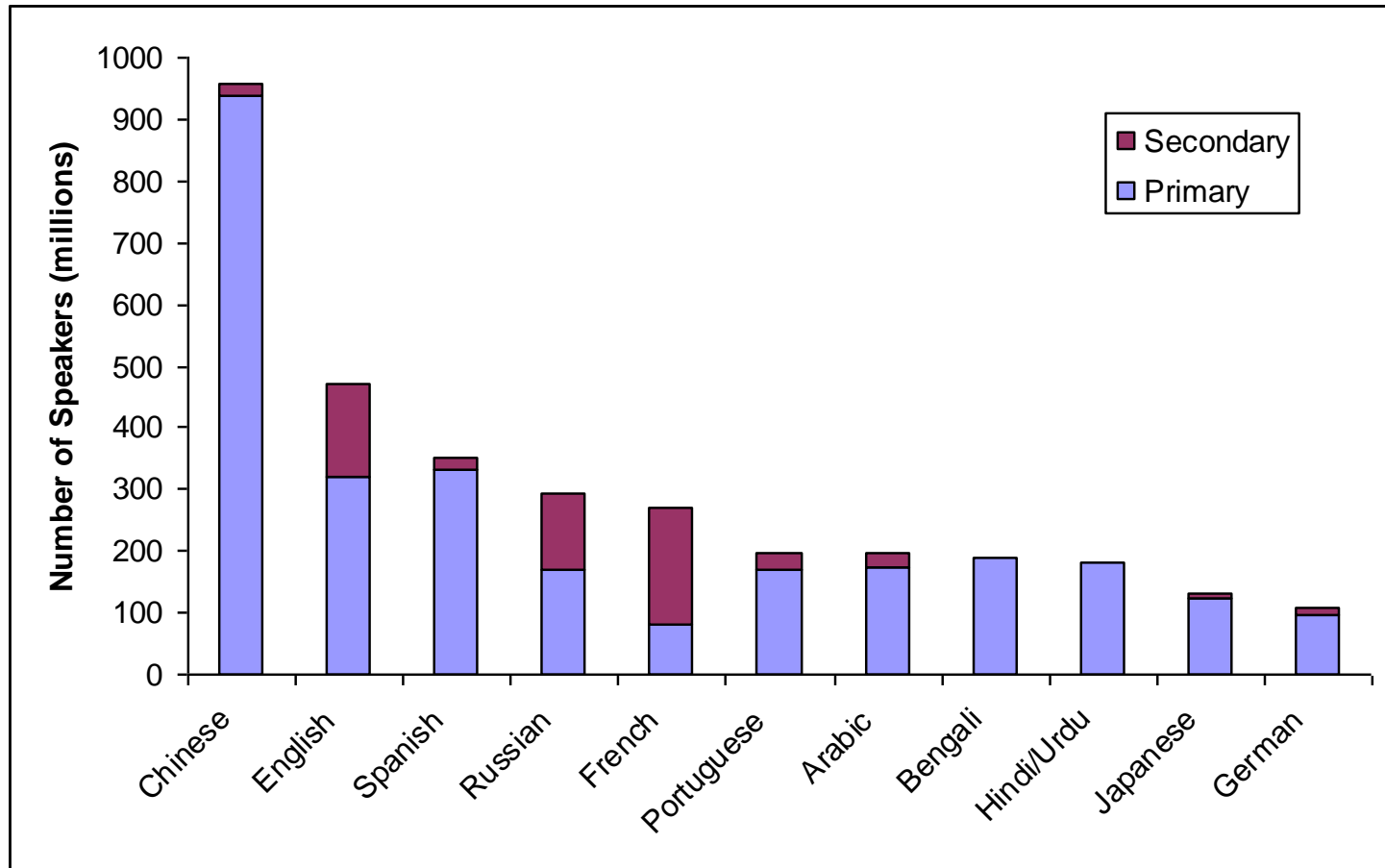
Link Structure of the Web

Global Internet Users

Web Pages

Legend:
- English
- Chinese
- Spanish
- Japanese
- Portuguese
- German
- Arabic
- French
- Russian
- Korean

Outer ring: 33%, 64%, 28%, 9%, 6%, 6%, 5%, 5%, 4%, 4%, 4%, 2%

Inner ring: 0%, 5%, 4%, 6%, 2%, 8%, 2%, 4%, 5%

Most Widely-Spoken Languages

Global Trade

Leading economies of merchandise trade, 2009

Imports (Billion USD) vs Exports (Billion USD)

Labeled countries: United States, China, Germany, France, Japan, Netherlands
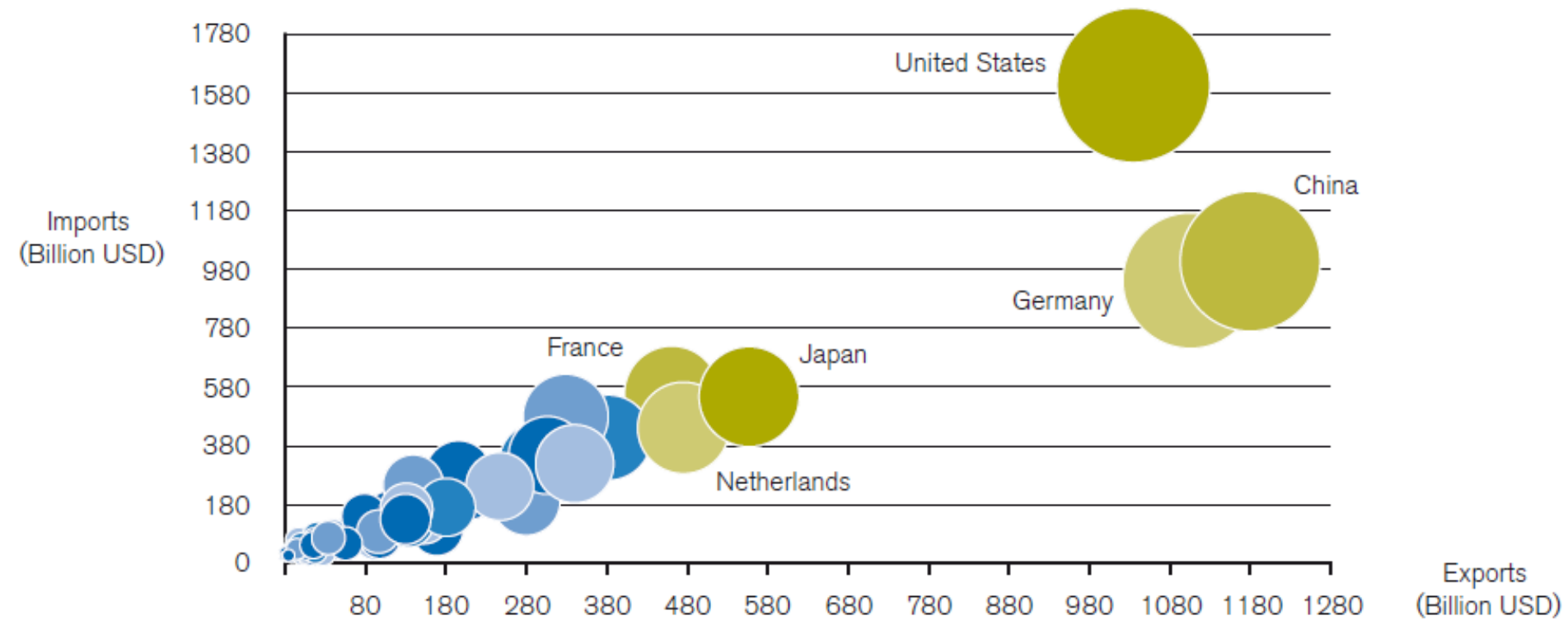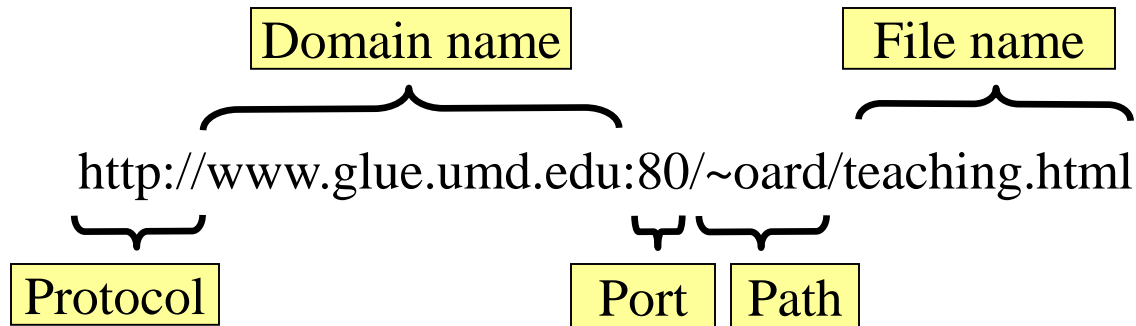
Why is there a Web?

- Affordable storage
  - 300,000 words/$ in 1995
- Adequate backbone capacity
  - 25,000 simultaneous transfers in 1995
- Adequate "last mile" bandwidth
  - 1 second/screen in 1995
- Display capability
  - 10% of US population in 1995
- Effective search capabilities
  - Lycos and Yahoo were started in 1995

Web Standards

- HTML
  - How to write and interpret the information

- URL
  - Where to find it

- HTTP
  - How to get it

Uniform Resource Locator (URL)

- Uniquely identify Web pages

Domain name · · · File name

http://www.glue.umd.edu:80/~oard/teaching.html

Protocol · · · Port · Path

HyperText Markup Language (HTML)

- Simple document structure language for Web

- Advantages
  - Adapts easily to different display capabilities
  - Widely available display software (browsers)

- Disadvantages
  - Does not directly control layout

"Hello World" HTML

```html
<html>
<head>
<title>Hello World!</title>
</head>

<body>

<p>Hello world! This is my first webpage!</p>

</body>
</html>
```

This is the actual content of the HTML document

Hands On:
Learning HTML From Examples

- Use Internet Explorer to find a page you like
  - http://www.glue.umd.edu/~oard

- On the "Page" menu select "View Source" (in IE)
  - Opens a notepad window with the source

- Compare HTML source with the Web page
  - Observe how each effect is achieved

Hands On: "Adopt" a Web Page

- Modify the HTML source using notepad
  - For example, change the page to yours

- Save the HTML source somewhere
  - In the "File" menu, select "Save As"
  - **Put the name in quotes (e.g.,** "test.html")

- FTP it to your ../pub directory on terpconnect

- View it
  - http://terpconnect.umd.edu/~(yourlogin)/test.html

Tips

- Edit files on your own machine
  - Upload when you're happy
- Save early, save often!
- Reload browser to see changes
- File naming
  - **<u>Don't use spaces</u>**
  - Punctuation matters

HTML Document Structure

- "Tags" mark structure
  - <html>a document</html>
  - an ordered list
  - <i>something in italics</i>
- Tag name in angle brackets <>
  - Not case sensitive (unlike XML)
- Open/Close pairs
  - Close tag is sometimes optional (unlike XML)

Logical Structure Tags

- Head
  - Title

- Body
  - Headers: <h1> <h2> <h3> <h4> <h5>
  - Lists: <ol>, <ul> (can be nested)
  - Paragraphs:<p>
  - Definitions: <dt><dd>
  - Tables: <table> <tr> <td> </td> </tr> </table>
  - Role: <cite>, <address>, <strong>, …

Physical Structure Tags

- Bold: \\

- Italics: \<i>\</i>

- Typeface: \\

- Size: \\

- Color: \\

(Hyper)Links

index.html

```
<html>
<head>
<title>Hello World!</title>
</head>
<body>
<p>Hello world! This is my first webpage!</p>
<p>Click <a href="test.html">here</a> for another page.</p>
</body>
</html>
```

test.html

```
<html>
<head>
<title>Another page</title>
</head>
<body>
<p>This is another page.</p>
</body>
</html>
```

Hypertext "Anchors"

- Internal anchors: somewhere on the same page
  - Students
    - Links to: Student Information

- External anchors: to another page
  - CLIS
  - CLIS students

- URL may be complete, or relative to current page
  - 2

- File name (in URL) is case sensitive (on Unix servers)
  - Protocol and domain name are not case sensitive

Images

- <img src="*URL*"> *or* <img src="*path/file*">
  - <img src="http://www.clis.umd.edu/IMAGES/head.gif">
  - SRC: can be url or path/file
  - ALT: a text string
  - ALIGN: position of the image
  - WIDTH and HEIGHT: size of the image
- Can use as anchor:
  - <img src=*URL2*>
- Example:
  - http://www.umiacs.umd.edu/~daqingd/Image-Alignment.html

Tables

eenie	mennie	miney
mo	catch	a tiger
by	the	toe

<table>

<tr><td> eenie **</td><td>** mennie **</td><td>** miney **</td>** **</tr>**

<tr><td> mo **</td><td>** catch **</td><td>** a tiger **</td>** **</tr>**

<tr><td> by **</td><td>** the **</td><td>** toe **</td>** **</tr>**

</table>

Table Example

```
<table align="center">
<caption align="right">The caption</caption>
    < tr align="LEFT">
        <th> Header1 </th>
        <th> Header2</th>
    </tr>
    <tr><td>first row, first item </td>
        <td>first row, second item</td></tr>
    < tr><td>second row, first item</td>
        <td>second row, second item</td></tr>
</table>
```

Wire Framing

- Design, brainstorming technique
  - Highly conceptual
  - Low-fi
  - Low risk, low cost
  - Rapid

- Incorporates three elements of Web design:
  - Information design: *Which information will go where?*
  - Navigation design: *How will users get around?*
  - Interface design: *How do elements convey functionality?*

GORMLESS THE FIRST

INTRO TEXT

More...

Gallery #1

Gallery #2

Gallery #3

More....

"My story"

"multimedia"

"become gormless"

BUTTON #1

BUTTON #2

BUTTON #3
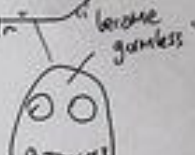
ME AT THE LOCAL SWIMMING POOL

LATEST VIDEO   MORE

CALL TO ACTION

FOOTER

INTERIOR: PRODUCT DETAIL PAGE v1

SAME WIDTH AS COLUMN

800 NUMBER

CHAT BADGE

data technologies inc.

Faxman SDK

INTRO BLOCK ON PRODUCT W/ HTML TYPE + LOGO GRAPHICS

INTRO TO BULLET ITEMS

BULLET FEATURES

CLOSING

CROSS-SELL FOR IMAGEMAN OR FAX BOARDS?
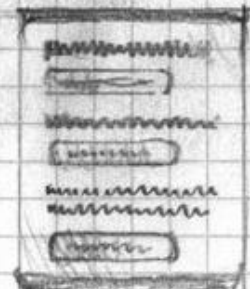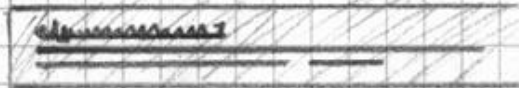
BUY MODULE. WHAT IF MOST EXPENSIVE IS @ THE TOP?

HELP + SUPPORT MODULE

LOGOS FOR APPROVAL - OR LOGOS OF COMPANIES WHO USE IT (CASE STUDIES)

WILSON SWORN IN AS PRESIDENT; PLEDGES HIMSELF TO JUSTICE; BIGGEST INAUGURAL THRONG

DUAL SUBWAY NOW ADOPTED

Greeting of Wilson and Taft.

OBAMA OFFERS LIBERAL VISION: 'WE MUST ACT'

Speech Gives Climate Goals Center Stage

By RICHARD W. STEVENSON
and JOHN M. BRODER

WASHINGTON — President Obama made addressing climate change the most prominent policy vow of his second Inaugural Address, setting in motion what Democrats say will be a deliberately paced but aggressive campaign built around the use of his executive powers to sidestep Congressional opposition.

Inaugural Stresses Theme of Civil and Gay Rights — Safety Net Praised

By PETER BAKER

WASHINGTON — Barack Hussein Obama ceremonially opened his second term on Monday with an assertive Inaugural Address that offered a robust articulation of modern liberalism in America, arguing that "preserving our individual freedoms ultimately requires collective action."

Feeling a Sense of Urgency To Leave His Mark
He did not say the words, but President Obama suffused his address with the spirit of a favorite phrase: the Rev. Dr. Martin Luther King Jr.'s call to heed "the fierce urgency of now." This was a man cognizant of his time left in office.
BY RICHARD W. STEVENSON, PAGE A4

A Diverse Crowd Celebrates A Moment in History
From the musicians who marched with a gay and lesbian band to high school mariachi performers from Texas to scores of elegant African-American women, the faces of Inauguration Day were the faces of those once left behind.
BY SHERYL GAY STOLBERG, PAGE A11

A First Lady Unafraid To Look Like a Million Bucks
Fashion is no longer the forbidden subject it once was in American politics. Embracing expensive designer clothes has not been a problem for Michelle Obama in the way that similar pursuits haunted Nancy Reagan, or even Jacqueline Kennedy.
BY ERIC WILSON, PAGE A6

Even if It Enrages Your Boss, Social Net Speech Is Protected

BY STEVEN GREENHOUSE

As Facebook and Twitter become as central to workplace conversation as the company cafeteria, federal regulators are ordering employers to scale back policies that limit what workers can say online.

Algeria Defends Tough Response to Hostage Crisis as Toll Rises

By ADAM NOSSITER
and ERIC SCHMITT

ALGIERS — The prime minister of Algeria offered an unapologetic defense on Monday of the country's tough actions to end the Sahara hostage crisis, saying that the militants who had carried out the kidnappings intended to kill all their captives and that the army saved many from death by attacking.

Rescue workers with the coffin of one of the slain hostages.

The New York Times

OBAMA OFFERS LIBERAL VISION: 'WE MUST ACT'

Speech Gives Climate Goals Center Stage

By RICHARD W. STEVENSON and JOHN M. BRODER

WASHINGTON — President Obama made addressing climate change the most prominent policy vow of his second Inaugural Address, setting in motion what Democrats say will be a deliberately paced but aggressive campaign built around the use of his executive powers to sidestep Congressional opposition.

"We will respond to the threat of climate change, knowing that failure to do so would betray our children and future generations," Mr. Obama said on Monday at the start of eight sentences on the subject, more than he devoted to any other specific area. "Some may still deny the overwhelming judgment of science, but none can avoid the devastating impact of raging fires, and crippling drought, and more powerful storms."

The central place he gave to the subject seemed to answer the question of whether he considered it a realistic second-term priority. He devoted scant attention to it in the campaign and has delivered a mixed message about its importance since the election.

Mr. Obama is heading into the effort having extensively studied the lessons from his first term, when he failed to win passage of comprehensive legislation to reduce emissions of the gases that cause global warming. This time, the White House plans to avoid such a fight and instead focus on what it can do administratively to reduce emissions from power plants, increase the efficiency of home appliances and have the federal government itself produce less carbon pollution.

Mr. Obama's path on global warming is a case study in his evolving sense of the limits of his power and his increased willingness to work around intense conservative opposition rather than seek compromise. After coming to office four years ago on a pledge to heal the planet and turn back the rise of the seas, he is proceeding cautiously this time, Democrats said, intent on making sure his approach is vetted politically, economically and technologically so as not to risk missing what many environmentalists advocates say could be the last best chance for years to address the problem.

The centerpiece will be action by the Environmental Protection
Continued on Page A13

Feeling a Sense of Urgency To Leave His Mark

He did not say the words, but President Obama suffused his address with the spirit of a favorite phrase: the Rev. Dr. Martin Luther King Jr.'s call to heed "the fierce urgency of now." This was a man cognizant of his time left in office.
BY RICHARD W. STEVENSON, PAGE A11

A Diverse Crowd Celebrates A Moment in History

From the musicians who marched with a gay and lesbian band to high-school mariachi performers from Texas to scores of elegant African American women, the faces of Inauguration Day were the faces of those once left behind.
BY SHERYL GAY STOLBERG, PAGE A11

A First Lady Unafraid To Look Like a Million Bucks

Fashion is no longer the forbidden subject it once was in American politics. Embracing expensive designer clothes has not been a problem for Michelle Obama in the way that similar pursuits haunted Nancy Reagan, or even Jacqueline Kennedy.
BY ERIC WILSON, PAGE A11

Inaugural Stresses Theme of Civil and Gay Rights — Safety Net Praised

By PETER BAKER

WASHINGTON — Barack Hussein Obama ceremonially opened his second term on Monday with an assertive Inaugural Address that offered a robust articulation of modern liberalism in America, arguing that "preserving our individual freedoms ultimately requires collective action."

On a day that echoed with refrains from the civil rights era and tributes to the Rev. Dr. Martin Luther King Jr., Mr. Obama dispensed with the post-partisan appeals of four years ago to lay out a forceful vision of advancing gay rights, showing more tolerance toward illegal immigrants, preserving the social welfare safety net and acting to stop climate change.

At times he used his speech, delivered from the West Front of the Capitol, to reprise arguments from the fall campaign, rebutting the notion expressed by conservative opponents that America risks becoming "a nation of takers" and extolling the value of proactive government in society. Instead of declaring the end of "petty grievances," as he did four years ago, he challenged Republicans to step back from their staunch opposition to his agenda.

"Progress does not compel us to settle centuries-old debates about the role of government for all time — but it does require us to act in our time," he said in the 18-minute address. "For now decisions are upon us, and we cannot afford delay. We cannot mistake absolutism for principle or substitute spectacle for politics or treat name-calling as reasoned debate. We must act."

Mr. Obama used Abraham Lincoln's Bible, as he did four years ago, but this time added Dr. King's Bible as well to mark the holiday honoring the civil rights leader. He became the first president ever to mention the word "gay" in an Inaugural Address as he equated the drive for same-sex marriage to the quests for racial and gender equality.

The festivities at the Capitol came a day after Mr. Obama officially took the oath as part of a ceremony with his family at the White House on the date set by the Constitution. With Inauguration Day falling on a Sunday, the swearing-in was then repeated for an energized mass audience a day later, accompanied by the pomp and parade that typically surround the quadrennial tradition.

Hundreds of thousands of people gathered on a brisk but bright day, a huge crowd by any measure, though far less than the record turnout four years ago. If the day felt restrained compared with the historic mood the last time, it reflected a more restrained moment in the life of the country. The hopes and expectations that loomed so large with Mr. Obama's taking the office in 2009, even amid economic crisis, have long since faded into a starker sense of the limits of his presidency.

Now lit and noticeably grayer, Mr. Obama appeared alternately upbeat and reflective. When he re-entered the Capitol at the conclusion of the ceremony, he stopped his entourage to turn back toward the cheering crowds on the National Mall.

"I want to take a look, one more time," he said. "I'm not going to see this again."

If the president won wistful, his message was firm. He largely eschewed foreign policy except to recommend engagement over war, and instead focused on addressing poverty and injustice at home. He did little to adopt the language of the opposition,
Continued on Page A12

Even if It Enrages Your Boss, Social Net Speech Is Protected

By STEVEN GREENHOUSE

As Facebook and Twitter become as central to workplace conversation as the company cafeteria, federal regulators are ordering employers to scale back policies that limit what workers can say online.

Employers often seek to discourage comments that paint them in a negative light. Don't discuss company matters publicly, a typical social media policy will say, and don't disparage the company itself. Violations can be a firing offense.

But in a series of recent rulings and advisories, labor regulators have declared many such blanket restrictions illegal. The National Labor Relations Board says workers have a right to discuss work conditions freely and without fear of retribution, whether
Continued on Page A3

Algeria Defends Tough Response to Hostage Crisis as Toll Rises

By ADAM NOSSITER and ERIC SCHMITT

ALGIERS — The prime minister of Algeria offered an unapologetic defense on Monday of the country's tough actions to end the Sahara hostage crisis, saying that the militants who had carried out the kidnappings intended to kill all their captives and that the army saved many from death by attacking.

But the assertion came as the death toll of foreign hostages rose sharply, to 37, and as American officials said they had offered sophisticated surveillance help that could minimize casualties, both before and during the military operation to retake a seized gas field complex in the Algerian desert.

At least some of the assistance was accepted, they said, but there were still questions about whether Algeria had taken all available steps to avert such a bloody outcome.

Caption: Rescue workers with the coffin of one of the slain hostages.

American counterterrorism officials and experts said they would have taken a more cautious approach, using detailed surveillance to gain an information-tion advantage and hopefully outmaneuver the militants. But others declined to second-guess the

Algerians, saying events had unfolded so rapidly that the government might have felt it had no choice but to kill the kidnappers, even if hostages died in the process.

The debate over how the Algerians handled one of the worst hostage-taking episodes in recent memory reflects conflicting ideas over how to manage such mass abductions in an age of suicidal terrorist acts in a post-9/11 world.

The Algerians — and some Western supporters — argue that the loss of innocent lives is unavoidable when confronting fanatics who will kill their captives anyway, while others say modern technology provides some means of minimizing the deaths.

At a news conference in Algiers, the prime minister, Abdelmalek Sellal, portrayed the military's deadly assaults on the Islamist militants who had stormed and occupied an internationally run gas-producing complex last Wednesday in remote eastern Algeria as a matter of national character and pride.

"The whole world has understood that the reaction was courageous," Mr. Sellal said, calling the abductions an attack "on the stability of Algeria."

"Algerians are not people who attack
Continued on Page A8

The New York Times

WORLD
U.S.
POLITICS
NEW YORK
BUSINESS
DEALBOOK
TECHNOLOGY
SPORTS
SCIENCE
HEALTH
ARTS
STYLE
OPINION

Autos
Blogs
Books
Cartoons
Classifieds
Crosswords
Dining & Wine
Education
Event Guide
Fashion & Style
Home & Garden
Jobs
Magazine
Movies
Music
Obituaries
Public Editor
Real Estate
Sunday Review
T Magazine
Television
Theater
Travel
Weddings / Celebrations

MULTIMEDIA
Interactives
Photography
Video

TOOL S & MORE
Alerts
Beta 620
Corrections
Mobile
Movie Tickets
Learning Network
Newsletters
NYT Events
NYT Store
Theater Tickets
Times Machine
TimesLimited
Times Skimmer

BREAKING NEWS 9:41 AM ET

Obama to Announce Return of 34,000 Troops From Afghanistan by 2014

North Korea Confirms It Conducted 3rd Nuclear Test

By DAVID E. SANGER and CHOE SANG-HUN

The test, estimated to be far larger than the country's previous two, was the first under its new leader, Kim Jong-un, and led to a crescendo of condemnation from leaders worldwide.

Post a Comment | Read (104)

NEWS ANALYSIS
A Big Challenge to China's New Leader

By JANE PERLEZ

Whether Xi Jinping gets tough on North Korea could tell the United States what kind of leader he will be.

Caption: Chung Sung-Jun/Getty Images
People watched a television broadcast reporting on North Korea's nuclear test at a railway station in Seoul, South Korea, on Tuesday.

RELATED COVERAGE
- Test Seen as Kim's Push for Credibility 5:56 AM ET
- Chinese React Online to Nuclear Test 8:20 AM ET
- Timeline: North Korea's Nuclear Program

Slower Growth of Health Costs Eases the Budget Deficit

By ANNIE LOWREY

The slowdown, which experts attributed to both a weak economy and structural changes, could help alleviate the country's long-term fiscal problems.

Post a Comment | Read (37)

WASHINGTON MEMO
Watching Obama for Signs of Change

By JACKIE CALMES

As the president readies his State of the Union address, many associates say he is exhibiting an assertiveness in contrast to the caution he showed in his first term.

- How the Union's State Got So 'Strong' Over Time

Militant Threats Test Role of a Pentagon Command in

NEWS ANALYSIS
A Turbulent Tenure for a Quiet Scholar

By LAURIE GOODSTEIN

When it came to the Roman Catholic Church's major challenges, Pope Benedict XVI often seemed to carom from one crisis to the next.

Caption: Pierre-Philippe Marcou/AFP — Getty

Benedict Successor Will Lead Church at Crossroads
By RACHEL DONADIO and ELISABETTA POVOLEDO
Conservatives who favor a smaller, more fervent church will confront those who feel the faith can grow its reach.

RELATED COVERAGE
- Vatican Tackles Logistics of Pope's Exit 8:44 AM ET
- Pope Was Fitted With Heart Pacemaker 7:14 AM ET
- GRAPHIC: The Catholic Church's Southward Shift

Airline Industry at Its Safest Since Dawn of Jet Age
By JAD MOUAWAD and CHRISTOPHER DREW
Four years without a fatal crash is an American record unmatched since propeller planes gave way to the jet.
- Boeing Battery Maker's Bind Is Mirrored by Japanese TV Show
Post a Comment | Read (25)

Fears of the Future Haunt Budding Afghan Strivers
By GRAHAM BOWLEY
Thousands of Afghans have built homes and careers on an influx of foreign money and are

The Opinion Pages

POPE BENEDICT XVI
James Martin: A papal resignation, and the future of the church.
Jason Berry: Pope Benedict can still right the wrongs.
John Patrick Shanley: Farewell to an uninspiring pope.
Carol Zaleski: Benedict's resignation reminds us of the church's need for true renewal.

OP-ED COLUMNISTS
- Brooks: Carpe Diem Nation
- Bruni: A Spritz of Power
- Nocera: The 'Die Hard' Quandary
- Cohen: Why 'Zero Dark Thirty' Works

MORE IN OPINION
- Editorials: Modernize Registration | Move Up NYC's Primaries
- Room for Debate: What to Ask the Treasury Nominee

MARKETS »   At 9:30 AM ET

	S&P 600	Dow	Nasdaq
	1,516.30	13,980.63	3,191.04
	−0.71	+9.29	−0.96
	−0.05%	+0.07%	−0.03%

GET QUOTES   My Portfolios »
Stock, ETFs, Funds
Go

SCIENCE TIMES

Straining to Hear and Fend Off Dementia

There is no clear explanation why, but compared with individuals with normal hearing, those with hearing loss have an increased risk of developing dementia.

A Match Made in Code

EHarmony, whose claims that its matchmaking algorithm can help people find their soul mates was criticized by academics, offers a defense.

Databases

- Database
  - Collection of data, organized to support access
  - Models some aspects of reality

- DataBase Management System (DBMS)
  - Software to create and access databases

- Relational Algebra
  - Special-purpose programming language

Structured Information

- Field   An "atomic" unit of data
  – number, string, true/false, …

- Record   A collection of related fields

- Table   A collection of related records
  – Each record is one row in the table
  – Each field is one column in the table

- Primary Key   The field that identifies a record
  – Values of a primary key must be unique

- Database   A collection of tables

A Simple Example

primary key

FITNESS CENTER DATABASE

Member file
Membership Plans file
Personal Trainer file
Training Session file

Membership Plans file

field names

Membership Code	Membership Name	Monthly Fee
L	Late Shift	$20.25
M	Matinee	$25.75
P	Prime Time	$45.50
R	Round the Clock	$55.50
W	Weekender	$39.50

records

fields

Characteristics of Membership Code field

Field Size	1
Format	
Input Mask	
Caption	Membership Code
Default Value	
Validation Rule	"L" Or "M" Or "P" Or "R" Or "W"
Validation Text	Please enter valid membership code.
Required	Yes
Allow Zero Length	No
Indexed	Yes (No Duplicates)

maximum characters in Membership Code field

Registrar Example

- Which students are in which courses?

- What do we need to know about the students?
  - first name, last name, email, department

- What do we need to know about the courses?
  - course ID, description, enrolled students, grades

A "Flat File" Solution

Student ID	Last Name	First Name	Department ID	Department	Course ID	Course description	Grades	email
1	Arrows	John	EE	EE	lbsc690	Information Technology	90	jarrows@wam
1	Arrows	John	EE	Elec Engin	ee750	Communication	95	ja_2002@yahoo
2	Peters	Kathy	HIST	HIST	lbsc690	Informatino Technology	95	kpeters2@wam
2	Peters	Kathy	HIST	history	hist405	American History	80	kpeters2@wma
3	Smith	Chris	HIST	history	hist405	American History	90	smith2002@glue
4	Smith	John	CLIS	Info Sci	lbsc690	Information Technology	98	js03@wam

Discussion Topic
Why is this a bad approach?

Goals of "Normalization"

- Save space
  - Save each fact only once

- More rapid updates
  - Every fact only needs to be updated once

- More rapid search
  - Finding something once is good enough

- Avoid inconsistency
  - Changing data once changes it everywhere

Relational Algebra

- Tables represent "relations"
    - Course, course description
    - Name, email address, department

- Named fields represent "attributes"

- Each row in the table is called a "tuple"
    - The order of the rows is not important

- Queries specify desired conditions
    - The DBMS then finds data that satisfies them

A Normalized Relational Database

Student Table

Student ID	Last Name	First Name	Department ID	email
1	Arrows	John	EE	jarrows@wam
2	Peters	Kathy	HIST	kpeters2@wam
3	Smith	Chris	HIST	smith2002@glue
4	Smith	John	CLIS	js03@wam

Department Table

Department ID	Department
EE	Electronic Engineering
HIST	History
CLIS	Information Stuides

Course Table

Course ID	Course Description
lbsc690	Information Technology
ee750	Communication
hist405	American History

Enrollment Table

Student ID	Course ID	Grades
1	lbsc690	90
1	ee750	95
2	lbsc690	95
2	hist405	80
3	hist405	90
4	lbsc690	98

Approaches to Normalization

- For simple problems (like the homework)
  - Start with "binary relationships"
    - Pairs of fields that are related
  - Group together wherever possible
  - Add keys where necessary

- For more complicated problems
  - Entity relationship modeling (LBSC 670)

Example of Join

Student Table

Student ID	Last Name	First Name	Department ID	email
1	Arrows	John	EE	jarrows@wam
2	Peters	Kathy	HIST	kpeters2@wam
3	Smith	Chris	HIST	smith2002@glue
4	Smith	John	CLIS	js03@wam

Department Table

Department ID	Department
EE	Electronic Engineering
HIST	History
CLIS	Information Stuides

"Joined" Table

Student ID	Last Name	First Name	Department ID	Department	email
1	Arrows	John	EE	Electronic Engineering	jarrows@wam
2	Peters	Kathy	HIST	History	kpeters2@wam
3	Smith	Chris	HIST	History	smith2002@glue
4	Smith	John	CLIS	Information Stuides	js03@wam

Problems with Join

- Data modeling for join is complex
  - Useful to start with E-R modeling

- Join are expensive to compute
  - Both in time and storage space

- But it is joins that make databases relational
  - Projection and restriction also used in flat files

Some Lingo

- "Primary Key" uniquely identifies a record
  - e.g. student ID in the student table

- "Compound" primary key
  - Synthesize a primary key with a combination of fields
  - e.g., Student ID + Course ID in the enrollment table

- "Foreign Key" is primary key in the <u>other</u> table
  - Note: it need not be unique in <u>this</u> table

Project

New Table

Student ID	Last Name	First Name	Department ID	Department	email
1	Arrows	John	EE	Electronic Engineering	jarrows@wam
2	Peters	Kathy	HIST	History	kpeters2@wam
3	Smith	Chris	HIST	History	smith2002@glue
4	Smith	John	CLIS	Information Stuides	js03@wam

SELECT Student ID, Department

Student ID	Department
1	Electronic Engineering
2	History
3	History
4	Information Stuides

Restrict

Student ID	Last Name	First Name	Department ID	Department	email
1	Arrows	John	EE	Electronic Engineering	jarrows@wam
2	Peters	Kathy	HIST	History	kpeters2@wam
3	Smith	Chris	HIST	History	smith2002@glue
4	Smith	John	CLIS	Information Stuides	js03@wam

WHERE Department ID = "HIST"

Student ID	Last Name	First Name	Department ID	Department	email
2	Peters	Kathy	HIST	History	kpeters2@wam
3	Smith	Chris	HIST	History	smith2002@glue

Entity-Relationship Diagrams

- Graphical visualization of the data model

- Entities are captured in boxes

- Relationships are captured using arrows

Registrar ER Diagram

Enrollment
Student
Course
Grade
…

has →

Student
Student ID
First name
Last name
Department
E-mail
…

has →

Course
Course ID
Course Name
…

associated with

Department
Department ID
Department Name
…
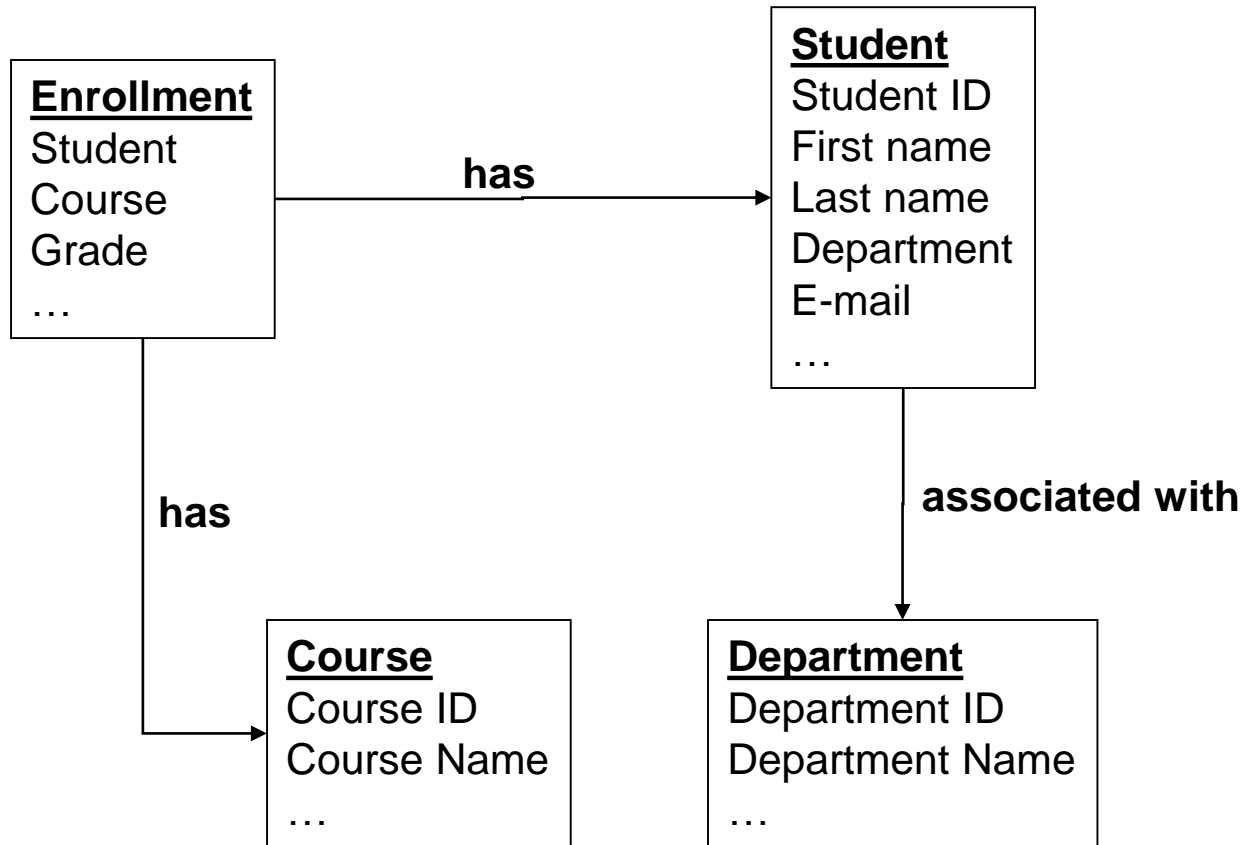
Getting Started with E-R Modeling

- What **questions** must you answer?

- What **data** is needed to generate the answers?
  - Entities
    - Attributes of those entities
  - Relationships
    - Nature of those relationships

- How will the user interact with the system?
  - Relating the question to the available data
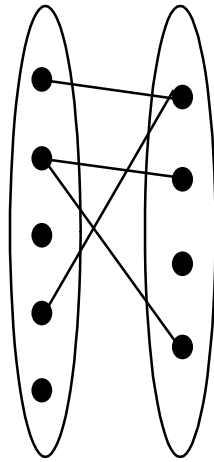 - Expressing the answer in a useful form
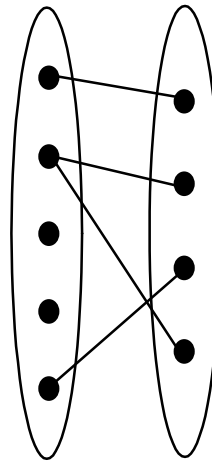
"Project Team" E-R Example
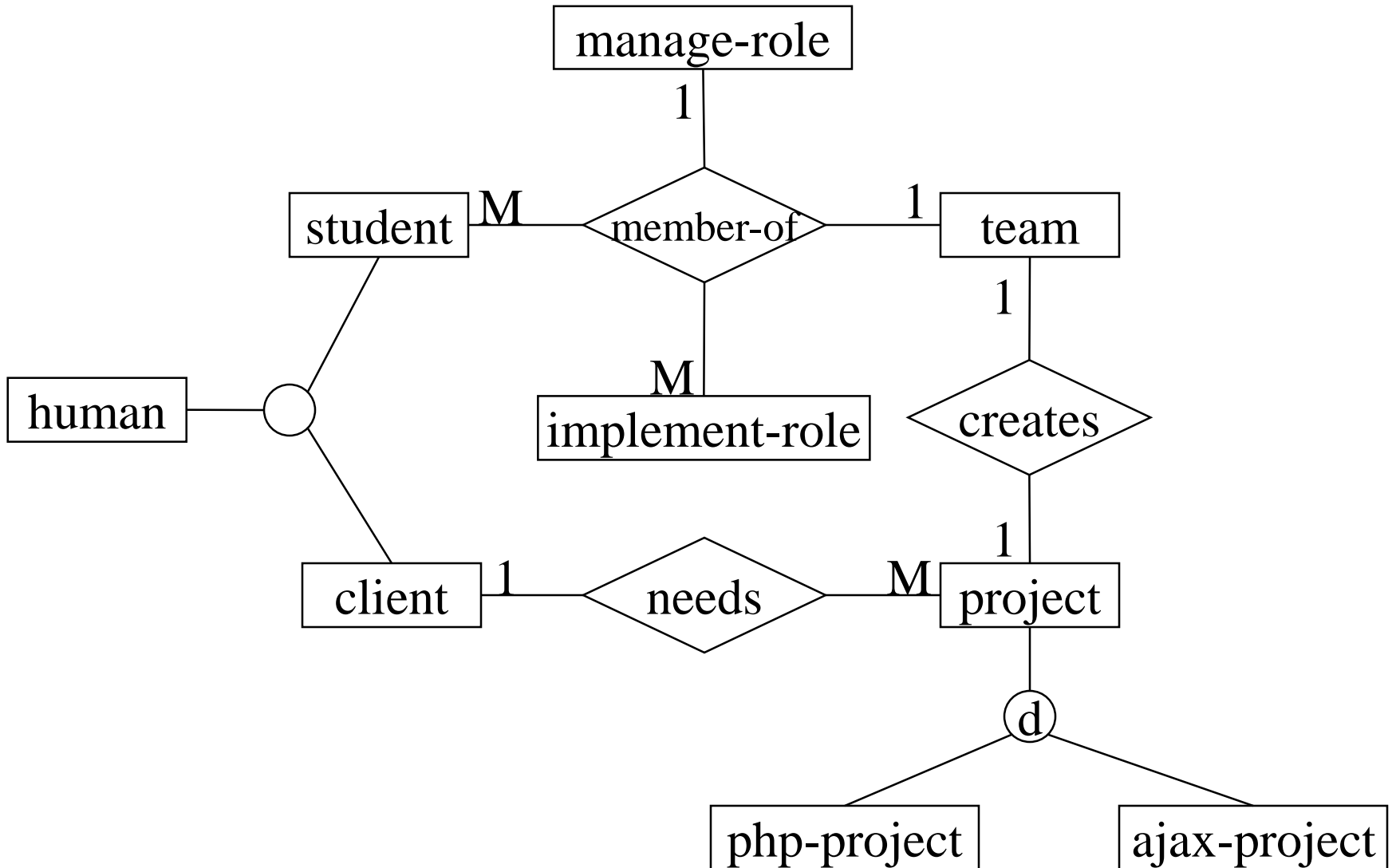
Components of E-R Diagrams

- Entities
  - Types
    - Subtypes (disjoint / overlapping)
  - Attributes
    - Mandatory / optional
  - Identifier
- Relationships
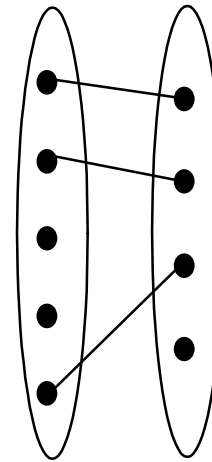 - Cardinality
  - Existence
  - Degree

Types of Relationships

Many-to-Many   **1-to-Many**   **1-to-1**

Making Tables from E-R Diagrams

- Pick a primary key for each entity
- Build the tables
  - One per entity
  - Plus one per M:M relationship
  - Choose terse but memorable table and field names
- Check for parsimonious representation
  - Relational "normalization"
  - Redundant storage of computable values
- Implement using a DBMS

Normalization

- 1NF: <u>Single-valued</u> <u>indivisible</u> (atomic) attributes
  - Split "Doug Oard" to two attributes as ("Doug", "Oard")
  - Model M:M implement-role relationship with a table

- 2NF: Attributes depend on <u>complete</u> primary key
  - (<u>id, impl-role</u>, name)->(<u>id</u>, name)+(<u>id, impl-role</u>)

- 3NF: Attributes depend <u>directly</u> on primary key
  - (<u>id</u>, addr, city, state, zip)->(<u>id</u>, addr, zip)+(<u>zip</u>, city, state)

- 4NF: Divide independent M:M tables
  - (<u>id</u>, role, courses) -> (<u>id</u>, role) + (<u>id</u>, courses)

- 5NF: Don't enumerate derivable combinations

Normalized Table Structure

- Persons: <u>id</u>, fname, lname, userid, password
- Contacts: id, ctype, cstring
- Ctlabels: <u>ctype</u>, string
- Students: <u>id</u>, team, mrole
- Iroles: <u>id, irole</u>
- Rlabels: <u>role</u>, string
- Projects: <u>team</u>, client, pstring

A More Complex ER Diagram

Cadastral Standard ERD
Version 2 - April 1999
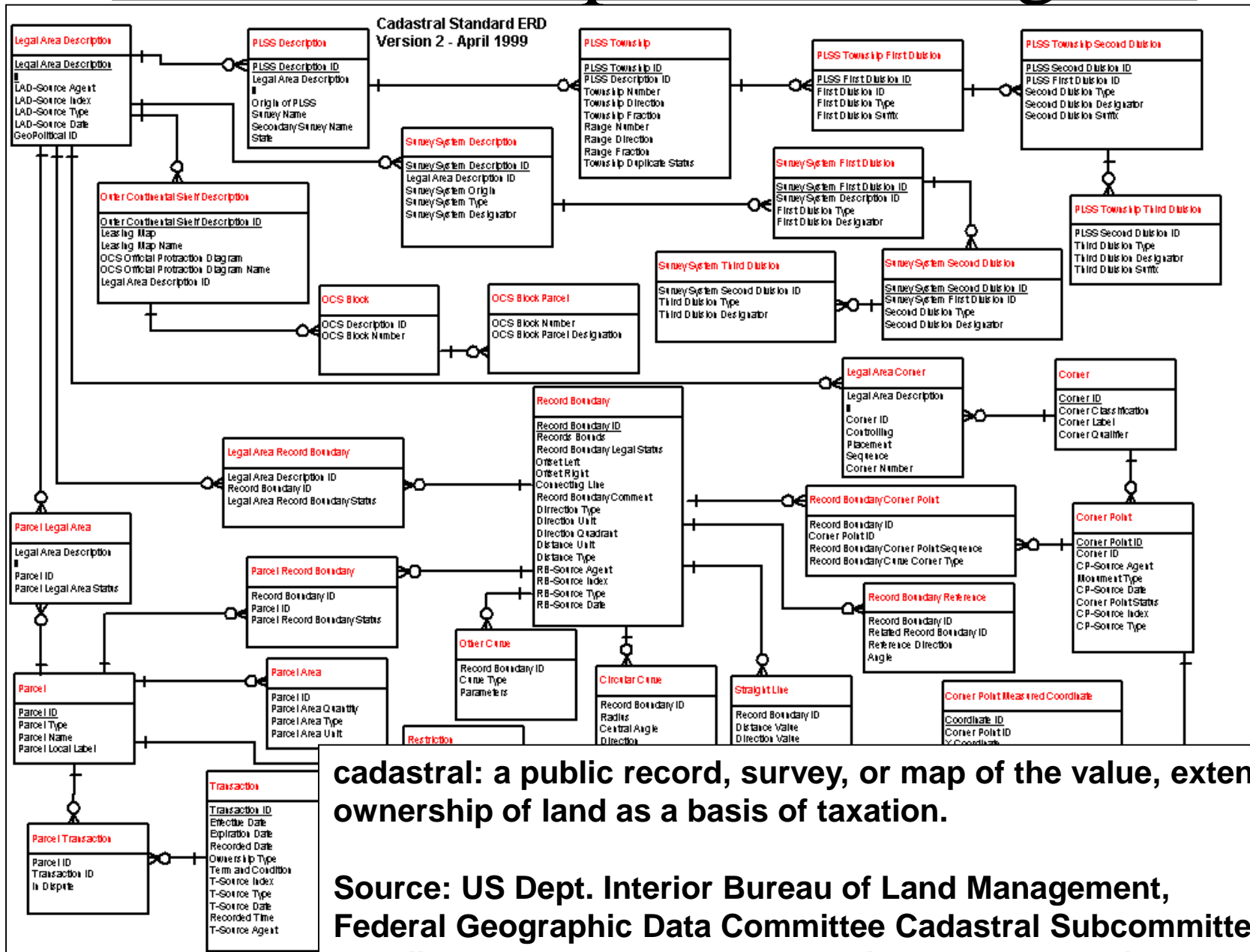
cadastral: a public record, survey, or map of the value, extent, and ownership of land as a basis of taxation.

Source: US Dept. Interior Bureau of Land Management,
Federal Geographic Data Committee Cadastral Subcommittee
http://www.fairview-industries.com/standardmodule/cad-erd.htm

Key Ideas

- Databases are a good choice when you have
  - Lots of data
  - A problem that contains inherent <u>relationships</u>

- Design before you implement
  - This is just another type of programming
  - The mythical person-month applies!

- Join is the most important concept
  - Project and restrict just remove undesired stuff

Before You Go

On a sheet of paper, answer the following (ungraded) question (no names, please):

What was the muddiest point in today's class?