#### Lecture

CS571 - Course Introduction

This content is protected and may not be shared, uploaded, or distributed.

#### General Information

- Lectures:
  - Session 30378D: Tue, Th 5:30PM 7:20PM (SGM 124)
  - Session 30016D: Tue, Th 5:30PM 7:20PM (DEN@Viterbi)
  - Session 30322D: Tue, Th 7:30PM 9:20PM (SGM 124)
- TAs: Zoom office hours on course website and Piazza
- Course website retired: https://www.csci571.com
- D2L Brightspace for all class materials, exams, grades, etc.: https://courses.uscden.net
- Assignments yes, attendance up to you
- Two Exams, one at half course and one towards the end
- Exams auto-graded using D2L Brightspace Quiz Tool
- Mobile Application final "comprehensive" project live demo
- Web storage: **GitHub Pages**
- Cloud Storage: Google Cloud, AWS or Azure

#### CS571: Web Technologies

- Instructor: Prof. Marco Papa
- Office hours, in-person and on Zoom:
  - Thursday 3:00PM-4:00PM PDT
  - In-person, in PHE 516
- E-mail: papa@usc.edu
- Lectures, live on Zoom and recorded, stored on D2L Brightspace, exclusively for DEN-enrolled students
- 24/7 Q&A access: Piazza, 30 min. average response time
- TA access: daily (Mon-Fri)
- Instructor access: weekly
- Quick way to ask a "personal" question: Private message to Instructors on Piazza

#### General Rules

## NO FACULTY D-CLEARANCE

Unless you are a "superstar" undergrad ©

# **COURSE OVERLAP**

Will allow, but you agree it is "your" problem

Fixed dates / times:

Exam #1: February 20, 5:30PM and 7:30PM PDT

Exam #2: April 25, 5:30PM and 7:30PM PDT

Final Mobile Project: May 1

Exams, "in-person", e-graded using D2L Brightspace Quiz Tool
Final Mobile project demo on Zoom video uploaded to D2L Brightspace

#### Benefits of Prof. Papa CSCI571

- Top 10 students in all sections combined offered available TA/Producer/Grader positions at end of semester. BY INVITATION ONLY
- CSCI571 Graders and Producers that are invited to MS Honors Program will be provided with Recommendation Letter (24 "recommended" students already awarded MS Honors)
- "Proof of skill" for H1B applications available at:
  - https://www.cs.usc.edu/skills-verification/
- Prof. Papa will act as "reference" in job / PhD applications for CSCI571 TAs, Course Producers and Graders

#### Learning Objectives

- By the end of the course, you should have acquired the following skills:
  - The ability to write RESTful API applications
  - The ability to set up Cloud services
  - The ability to design and code back-end scripts in Python and JavaScript
  - The ability to design and code front-end Web Applications
  - The ability to design and code Mobile Apps in Swift/SwiftUI or Java/Kotlin
  - The ability to design and code web front-end asynchronous applications using AJAX
  - The ability to design and code responsive web apps
  - The ability to read Web Services API documentation and use it in building Web applications
  - The ability to write Microservices and Containers
  - The ability to use a NoSQL database like MongoDB Atlas

#### Prerequisites

- The course is recommended for students with <u>deep</u> knowledge of at least two programming languages (JavaScript, Java, Python, C#, C++)
- CSCI 571 is **not a "beginner course"** for the Web; **it** is an "advanced course"
- If you are interested in beginner courses, check out these ITP courses:
  - ITP 104, "Introduction to Web Development".
  - ITP 301, "Front-End Web Development"
  - ITP 303, "Full-Stack Web Development"
  - ITP 304, "Back-End Web development"
  - ITP 341, "Android App Development"
  - ITP 342, "iOS App Development"
- In CSCI 571 you have <u>all of the above in 1 course!</u>
- Understand that the course load will be massive
- You need to have the time available to be successful
- Consider yourself warned

#### Course Technologies

- This course focuses on the phenomenon known as the World Wide Web
- Core technologies are:
  - HyperText Markup Language (HTML) and Cascading Style Sheets (CSS)
  - HyperText Transfer Protocol (HTTP)
  - Web servers, their configuration and performance properties
  - Server-Side programming using JavaScript and Python
  - Client-side programming using JavaScript and JS Frameworks
  - Ajax Development Style
- Newer Technologies of Interest
  - Responsive Website Design (Bootstrap, etc.)
  - JS Frameworks (Angular, React and Node.js)
  - Web Services (REST)
  - Web security, TOR, Dark web
  - Native Mobile frameworks (Java / Android and Swift / iOS)
  - React (native)
  - Cloud computing (AWS, GCP, Azure)
  - Serverless Applications, Containers, Docker
  - AWS Lambda, Google Cloud Functions, Azure Functions

#### Software and Storage

- Student Disk space on GitHub Pages:
  - https://pages.github.com/
  - GitHub Student Developer Pack: https://education.github.com/pack
  - Allows GitHub Pages with "private" repository
  - 1GB of free web space
  - Used for Assignment 1 and Table of Assignments
- Website / Web Services in the cloud
  - Amazon's Elastic Compute Cloud (AWS)
  - Google Cloud Platform (GCP) [recommended]
  - Microsoft Azure
  - AWS Lambda, GCP Functions, Azure Functions
  - Serverless.com
  - Node.js
  - MongoDB Atlas

#### Other Issues

- Piazza class news group
  - Activate your membership by self-joining at: piazza.com/usc/spring2024/csci571
  - Class Access Code: lafc3252usc
  - Discussions limited to course materials, assignments and exams
- Academic Integrity Policy
  - Do NOT submit the same program; you can discuss the project with fellow students, but do not develop code with other students; do not download code online; do not post code online; we use MOSS to check for plagiarism. We scan all the exams. See "Academic Integrity Policy".
- D2L Brightspace main site for course materials
  - Course slides (PDF)
  - Exams (Quiz Tool)
  - Assignments (Dropbox Folders)
  - Grades

# **BRIGHTSPACE**

#### Student Evaluations

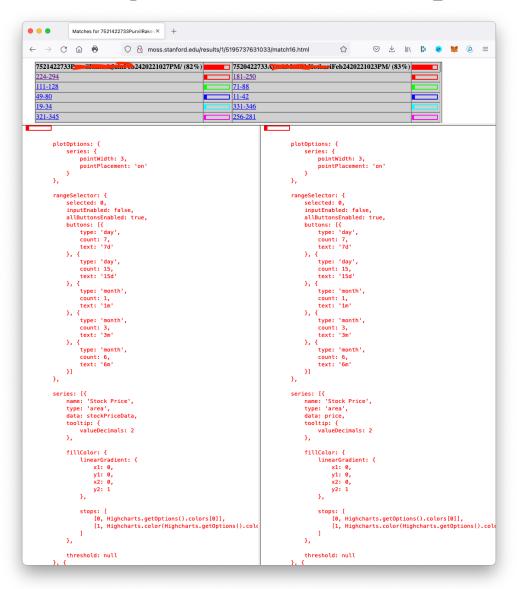
#### • Comments:

- "Amazing assignments. Learnt a lot on the course."
- "Projects seemed similar to an actual client for web development would ask for."
- "Even though the assignments were hard I learnt a lot from them."
- "It is not a fair game for beginners. I've spent almost 3 weeks to do a homework, and I still can not finish it on time."
- "Course projects are impressive!"
- "This class has posted assignments easily x10 times larger than other classes."
- "last two assignments take \*forever\*."
- "The homework assignments are so difficult."
- "I had to do so much googling on my own to learn about concepts used in the homework assignments."
- "Tough class with a lot of valuable assignments."
- "Massive assignments."
- "The workload of this course is too much, especially the last two homework."

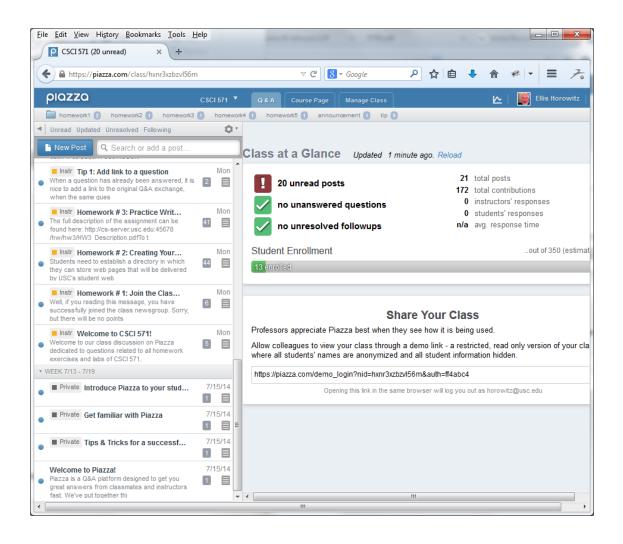
#### Academic Integrity Violations

```
Spring 2018 violations (16):
    - Sanctioned: 16, Appeals Panel: 8, Appeals to Dean: 2: no changes
    - F in course: 12, 0 + full letter grade reduction: 4 (C-, C-, C-, B-)
 Fall 2018 violations (0):
    - None!
 Spring 2019 violations (4):
    - F in course: 1
    - 0 + full letter grade reduction: 1 (B-), C in the course: 2
 Fall 2019 violations (6):
    - F in course: 1
    - 0 + full letter grade reduction: 5 (B-, C)
 Spring 2020 violations (4):
    - 0 + full letter grade reduction: 1 (A-, C)
 Spring 2021 violations (2)
    - Full letter grade reduction
 Spring 2022 violations (52)
    - F in course: 2, 0 in assignments, C, and D in course
 Fall 2022 violations (5)
    - 0 in assignments, F in course: 1, point deductions
 Spring 2023 violations (4)
   - 0 in assignments
• Fall 2023 violations (4)
    - 0 in assignments
```

## MOSS Plagiarism discovery tool



#### Piazza



#### Who am I?

- PhD in CS from USC, class of '88
- Initial career: MS Windows, Commodore Amiga developer
- System Architect -> Team Lead -> IT Project Manager -> VP Engineering
- CTO (Chief Technology Officer) at Luckman Interactive and CareerBuilder
- Part-time Faculty at USC 2003-2022
- Full-time Faculty at USC starting on August 2022
- Chief Technologist at LASC (Los Angeles Superior Court) 2002-2022
- Microsoft 365 SharePoint Project Manager 2014-2022
- Active member of LA CTO Forum (Silicon Beach CTOs)
- Season ticket holder of USC Football and 🔣 (Los Angeles Football Club) and supporter

#### Characterizing Web Content

There are very few studies that examine the types of content on the web, however . . .

(From IEEE Spectrum, Jan. 2004, pp. 75):

- Claim: 30% of the web is porn
- Claim: 30% of the web is duplicate information
- 50,000,000 pages are either new or changed each day
- 65% of the web pages are in English

(From Personal Computer World, Optenet, Sep. 2008):

- Claim: 35% of the web is porn, 11% is e-commerce
- http://www.optenet.com/en-us/new.asp?id=162 (N/A)

(From Forbes, Sept. 2011):

- Claim: 4% is porn, 13% are porn Web Searches
- http://www.forbes.com/sites/julieruvolo/2011/09/07/how-much-of-the-internet-is-actually-for-porn/

(From BBC, July 2013):

- Claim: is porn 4% or 37%?
- http://www.bbc.com/news/technology-23030090/

#### Sample Web Site (Small size)

- Here is a popular strategy for maintaining a web site
- The facts:
  - www.autobytel.com, new and used cars (now AutoWeb)
  - Market Cap: \$33.92M (Dec. 2019)
  - Quarterly Revenues: \$28.6M (3<sup>rd</sup> Quarter 2019)
  - Lead traffic: 31.7M visits (3<sup>rd</sup> Ouarter 2019)
  - Mobile version launched in 2012
  - Stock symbol: **AUTO** (Nasdag) delisted
- Original Microsoft solution:
  - Microsoft Windows Server
  - Microsoft IIS 7.5 web server
  - Microsoft SQL server database
  - Akamai CDN
- Today:
  - Microsoft Azure hosting
  - F5 BIG-IP OS, BigIP web server

#### Sample Web Site (Medium Size)

- Running a web site can get complicated; here is one example.
- The facts:
  - https://www.on-running.com, online sale of shoes and clothing goods
  - Public company (ONON on NYSE)
  - Market Cap: \$8.5B
  - Revenues: \$1.2B (2022)
  - 10.3M page visits per month
  - 5.69 average pages per visit
- The solution:
  - Cloudflare CDN hosting, DNS, SSL/TLS certs
  - Linux OS
  - cloudflare-nginx web server
  - Gzip compression, HSTS, HTML5

#### Sample Web Site (large size)

#### • The facts:

- www.etrade.com, online investing services and resources
- Market Cap: \$10.37B (Dec. 2019)
- Yearly Revenues: \$2.9B (12/2018)
- 60 million-page views per month
- average of 53,000 unique visitors per day
- 4.9 million accounts (Jan. 2015)
- 25,000 new retail accounts opened (Oct 2015)
- 1,952,000 customer transactions per month
- Taken over by Morgan Stanley in 2020 for \$13 billion

#### • The solution:

- AWS hosting
- IBM 90 xSeries running Linux/Citrix Netscaler, Apache and Tomcat web servers, AWS Route 53 (DNS)
- Hardware facility for load balancing and redundancy
- Oracle database system
- Proprietary programming systems

#### Web Server Farms

- Until recently all serious web sites were maintained using web server farms;
  - A group of computers acting as servers and housed in a single location;
  - Internet Service Providers (ISP's) provide web hosting services using a web server farm
- Hardware and software is used to load balance requests across the machines
- Other issues addressed by web server farms include:
  - Redundancy eliminates single point of failure; backup and failover strategy is required
  - **Security**, secure areas are placed behind firewalls which monitor web traffic, network address translation, port translation, SSL

#### Popular Web Hosting Services

- For individuals and small business:
  - 1&1

```
https://www.ionos.com
```

- GoDaddy.com

```
https://www.godaddy.com/hosting/web-hosting
```

- Yahoo

```
http://www.iwebhostingplans.com/yahoo/yahoowebhosting.as
```

- For companies willing to pay MUCH higher costs:
  - Rackspace

```
https://www.rackspace.com/?CMP=Google hosting
```

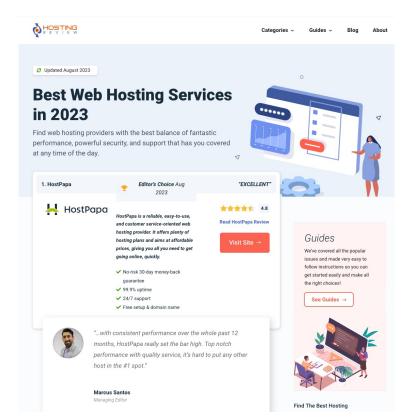
Network Solutions

```
http://www.networksolutions.com/hosting/
```

- Reviews and price comparisons:
  - https://www.hosting-review.com/top-ten-list/
    See next slide "monthly" prices
  - http://www.pcmag.com/categories/web-hosting

#### Web Hosting Services





#### Cloud Computing

- Cloud computing is Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand, like the electricity grid.
- Users no longer have need for expertise in, or control over, the technology infrastructure "in the cloud" that supports them.
- It typically includes web-based tools or applications that users can access and use through a web browser as if it were a program installed locally on their own computer.
- Typical cloud computing providers deliver common business applications online that are accessed from another Web service or software like a Web browser, while the software and data are stored on servers.
- The major cloud service providers include Amazon, Google, Microsoft, Salesforce, Skytap, HP, IBM, Amazon, Google and Apple (iCloud).

# An Example - Amazon's Elastic Compute Cloud

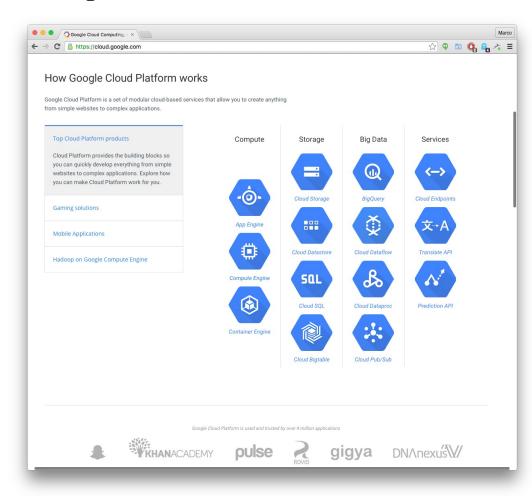
- A web service providing resizable compute capacity
- The "elastic" nature means the service instantly scales to meet demand with no up-front investment
- Users create an Amazon Machine Image (AMI), a virtual computer running your selected operating system (Linux, Windows, etc)
- Users use Amazon's Simple Storage Service (S3) for largescale, persistent storage
- You only pay for running AMI
- All accounts are limited to 5
   Elastic IPv4 addresses per region
- See: aws.amazon.com/ec2



Amazon currently runs in 8 regions: US East, US West (Oregon), US West (Northern CA), Ireland, Asia Pacific (Singapore), Asia Pacific (Tokyo), Asia Pacific (Sydney), South America (Sao Paulo)

#### An Example - Google Cloud Platform

- A web service providing basic Compute, Storage, Big Data and Services.
- Additional services for massively scalable Gaming solutions, Mobile Applications backend, and Apache Hadoop.
- App Engine A platform for building scalable web applications and mobile backends. App Engine scales applications automatically in response to the amount of traffic it receives.
- Compute Engine Offers predefined virtual machine configurations:Debian, CentOS, CoreOS, SUSE, Ubuntu, Red Hat, FreeBSD, or Windows 2008/2012.



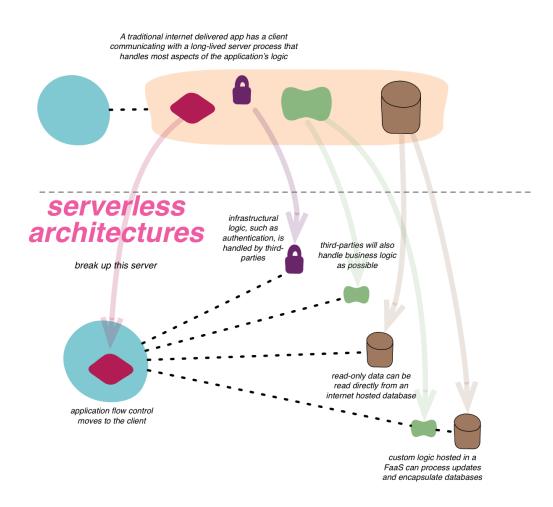
Google uses software-defined networking technology to route packets across the globe and enable fast edge-caching so that data is where it needs to be to serve users.

#### Serverless Architecture

- Internet based systems where the application development does not use the usual server process.
- They rely solely on a combination of:
  - third-party services, or Backend as a Service (BaaS)
  - client-side logic
  - service hosted remote procedure calls, or Function as a Service (FaaS).
- AWS Lambda is one of the most popular implementations of FaaS at present, but there are others. See:

https://aws.amazon.com/lambda/

• Serverless and contains will be covered later in the course



#### USC Has Many Web Servers Running







#### Hostnames matching usc.edu

▶ Q Search with another pattern?

#### 120 results (showing 1 to 20)

Rank	Site	First seen	Netblock	os	Site Report
72414	login.usc.edu ௴	March 2020	Amazon Technologies Inc.	Linux	B
79277	bytes.usc.edu &	September 2017	University of Southern California Linux - Ubuntu		B
93270	shibboleth. <b>usc.edu</b> 🗷	Febuary 2007	University of Southern California	unknown	
104928	mychart.m <b>usc.edu</b> ළ	June 2012	Medical University of South Carolina	Citrix NetScaler	B
143668	libguides.usc.edu &	August 2018	Amazon.com, Inc.	Linux	B
150569	viterbigradadmission. <b>usc.edu</b> 전	March 2017	WPEngine, Inc.	Linux	
152228	gradadm.usc.edu &	October 2017	Google LLC	Linux	B
155572	www.usc.edu &	August 1995	Amazon.com, Inc.	Linux	
160240	camel2.usc.edu &	January 2004	University of Southern California	unknown	B
193115	my.usc.edu ජ	April 2004	University of Southern California	unknown	B
196005	catalogue. <b>usc.edu</b> 🗹	April 2013	Amazon Technologies Inc.	Linux	
207538	www-scf.usc.edu &	August 1995	University of Southern California	unknown	B
219979	dornsife.usc.edu ど	May 2011	University of Southern California	unknown	B
242432	itservices.usc.edu &	April 2014	University of Southern California	unknown	

- Netcraft lists 120
   separate sites / web
   servers with usc.edu in
   their name, e.g.
- www.usc.edu
- mat.usc.edu
- www.cs.usc.edu
- dornsife.usc.edu
- web-applusc.edu
- www-scf.usc.edu
- However, some may not be connected to USC, e.g.
- www.usc.edu.au
- Check at:

https://searchdns.netcraft
.com

#### Web Browsers Use Standard Layout Engines

- WebKit is a software component used to render web pages; it is open source.
  - It is used by Google's Chrome and Apple's Safari web browsers
  - WebKit is also the name of the macOS system framework version of the engine that's used by Safari, Dashboard, Mail, and many other macOS applications
- **Gecko** is a layout engine developed by Mozilla Corporation, known as the layout engine of the Firefox web browser.
  - It is used to display web pages and, in some cases, an application's user interface.
  - It offers a rich programming API that makes it suitable for a wide variety of roles in Internet-enabled applications, such as web browsers
  - Its development originated with Netscape Communications Corporation
- **Chromium** is web browser project, developed by Google.
  - Engine originally based on Blink, a fork of Apple WebKit
  - Used in Chrome, Microsoft Edge, Opera, Samsung Internet and many others
- Some web kits and the browsers that use them:
  - Gecko-based: FireFox (Mozilla), Flock, Netscape
  - Trident-shells: Internet Explorer (Microsoft, <u>retired</u>), Edge (retired in <u>Aug. 2021</u>)
  - WebKit-based: Midori, Safari and Mobile Safari (Apple), Symbian<sup>3</sup> (Nokia) and many others
  - **Chromium-based**: Chrome, Microsoft Edge (2021+)

#### Capabilities of a Browser

- Web browsers fetch and display documents from other WWW sites; their capabilities include:
  - A mouse-driven graphical user interface
  - Display of
    - Hypertext documents conforming to latest HTML standard
    - Text with fonts, styles, and varying point sizes
    - Foreign-language character sets conforming to ISO-8859
    - Forms composed of edit boxes, check boxes, radio boxes, lists, text areas, etc.
    - Graphics in different formats (GIF, JPEG, MPEG, PNG, XBM) including monochrome, color

GIF = graphic interchange format, MPEG = Motion Picture Experts Group, JPEG = Joint Photographic Experts Group, PNG = Portable Network Graphics, XBM = x bitmap

#### Capabilities of a Browser



- Ability to invoke helper applications and plug-ins, (Obsoleted in HTML5) e.g.
  - Adobe Acrobat used to view pdf files
  - Windows Media Player to play digital sound files
  - Adobe Flash Player, used to display video. Retired in 2020.)
- Ability to communicate over a secure channel, using SSL
- Ability to maintain and exchange digital certificates
- Ability to run scripts in JavaScript
- Ability to run Java applets
  and Active X components (also
  obsoleted in HTML5)

#### The Browser Wars - Desktop Statistics

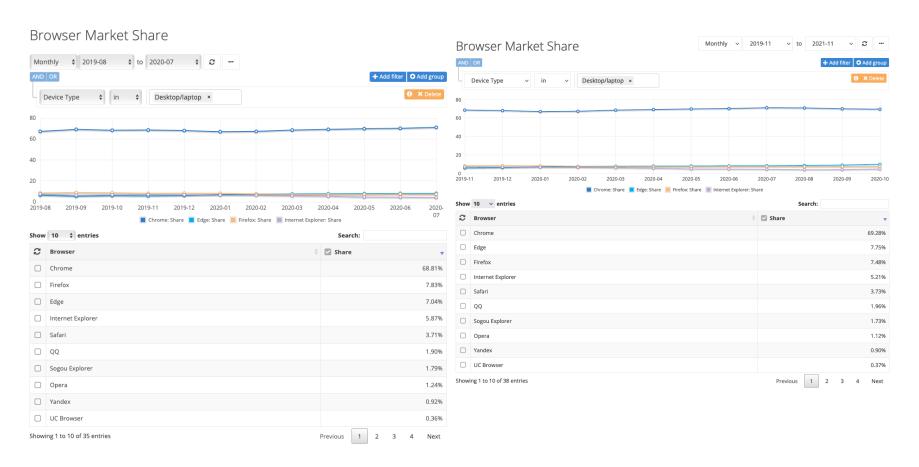
2023	<u>Chrome</u>	<u>Edge</u>	<u>Firefox</u>	<u>Safari</u>	<u>Opera</u>
November	77.4 %	10.6 %	4.9 %	3.9 %	2.4 %
October	78.0 %	10.3 %	4.8 %	3.9 %	2.3 %
September	78.8 %	10.3 %	4.6 %	3.4 %	2.2 %
August	80.1 %	9.8 %	4.6 %	3.0 %	1.8 %
July	79.9 %	10.0 %	4.8 %	3.0 %	1.8 %
June	79.8 %	9.4 %	4.3 %	3.5 %	2.1 %
May	79.4 %	9.3 %	4.5 %	3.7 %	2.3 %
April	79.5 %	9.0 %	4.6 %	3.8 %	2.2 %
March	79.6 %	8.8 %	4.7 %	3.7 %	2.3 %
February	79.7 %	8.6 %	4.8 %	3.9 %	2.2 %
January	79.7 %	8.2 %	5.0 %	3.9 %	2.3 %
2022	Chrome	Edge	Firefox	Safari	Opera
December	80.3 %	7.8 %	4.9 %	3.7 %	2.4 %
November	79.9 %	8.1 %	4.9 %	3.9 %	2.2 %

Conclusion of the above
study:

- Chrome is the clear winner
- Edge comes second, with Chrome engine
- Firefox next, but losing ground
- Safari and Opera having small percentages
- WebKit-related total over 90%

http://www.w3schools.com/browsers/browsers\_stats.asp

#### Desktop/laptop Browser Market Share Statistics



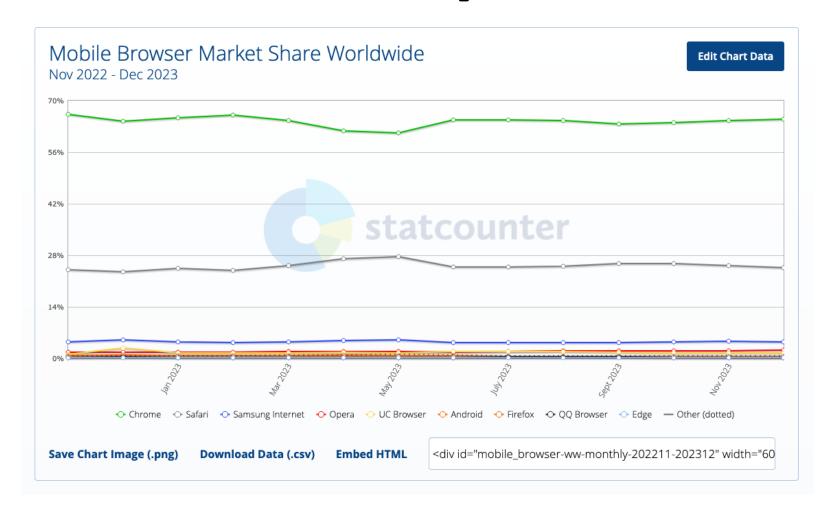
August 2020

December 2021

http://www.netmarketshare.com/ [retired]

Chrome leads with 69% market share.

## The Browser Wars Comparison (cont'd)



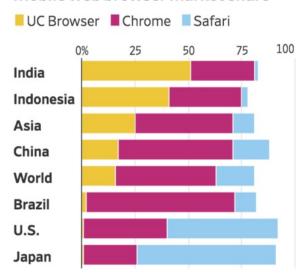
StatCounter Global Stats, Dec. 2022 - Dec. 2023, See http://gs.statcounter.com Chrome has the lead with about 65%, followed by Safari at 24%, Samsung, Opera and UC Browser from UCWeb of Alibaba Group of China.

## The Browser Wars Comparison (cont'd)

#### **Browser for the Next Billion**

Alibaba's mobile browser, UC Browser, has a larger market share than Google's Chrome in India and Indonesia, where many of the world's 'next billion users' are getting online for the first time.

#### Mobile web browser market share

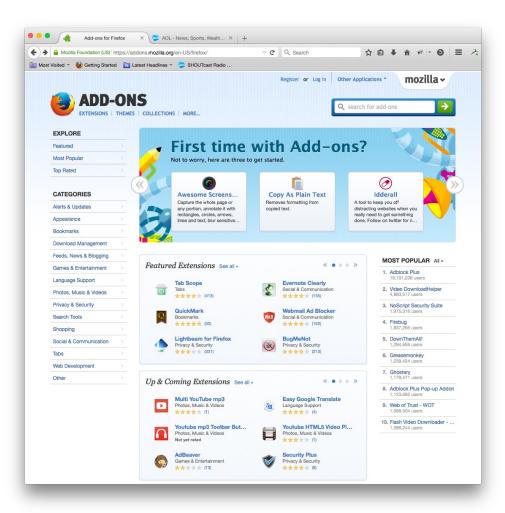


Note: Data Oct.-Dec. 2016 through Oct.-Dec. 2017.

Source: StatCounter

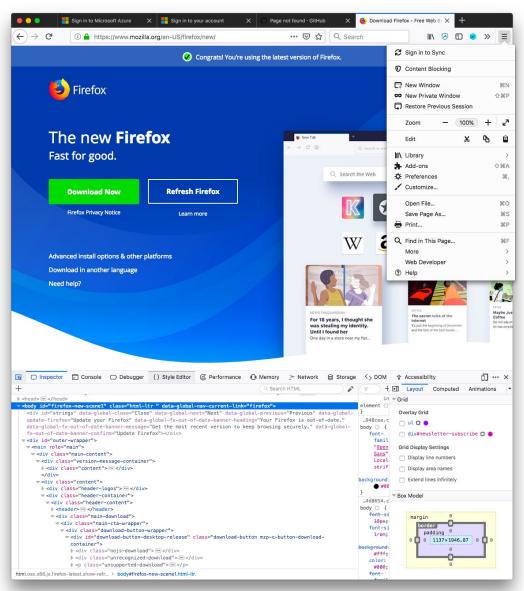
StatCounter Global Stats, Oct.-Dec. 2016 through Oct.-Dec. 2017, See https://www.wsj.com/articles/a-browser-youve-never-heard-of-is-dethroning-google-in-asia-1514808002

#### Browsers Have Many Plugins Available

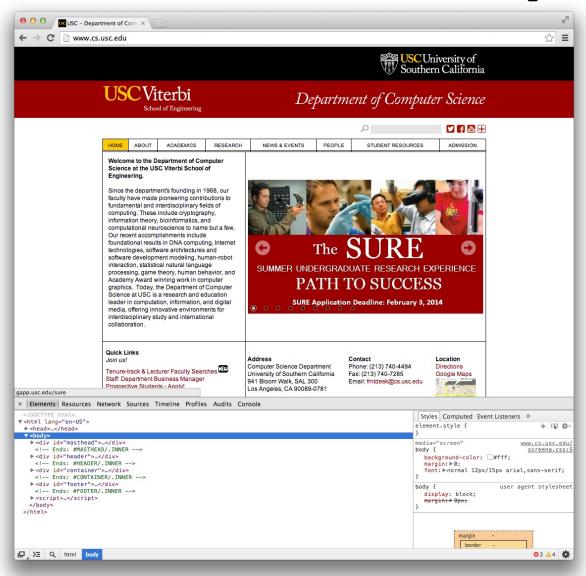


- Plug-ins are also listed as "Extensions"
- Firefox plug-ins that will be especially useful in this course are:
  - HTTP Header Live
  - https://addons.mozilla.org/en-US/firefox/addon/http-header-live/
- HTML5 does away with most video / audio plug-ins

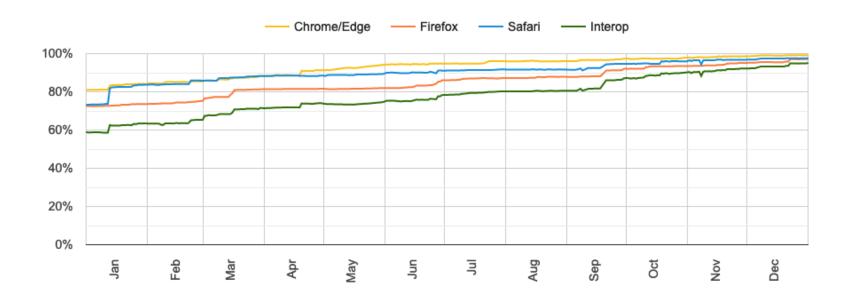
## Firefox: Tools | Web Developer



#### Chrome: Menu| More Tools | Developer Tools



## Browsers Interoperability



https://wpt.fyi/interop-2023

### Evolution of Web Sites

Client- centric Static	Server Applications  Databases  Dynamic web pages	Web services Multiple layers  Business and service Integration	Service Oriented Arch. (SOA) Client- centric	Multi- platform (desktop, tablet, phone) Client- centric	IoT, Wearables, Cloud computing, Serverless Arch. (BaaS, FaaS) Docker	Web Assembly, WebGPU, SQLite, File System Access API, Passkeys,
HTML Scripts CGI	ODBC, JDBC ASP Applets, ActiveX	XML, WML, SQL, .NET COM+, Beans	Ajax, Web 2.0, JSON	HTML5, CSS3, JS gestures	JS Frameworks AWS, GCP, Azure Microservices containers	Web Transitions API, Badging API
1 <sup>st</sup> gen	2 <sup>nd</sup> gen	3 <sup>rd</sup> gen	4 <sup>th</sup> gen	5 <sup>th</sup> gen	6 <sup>th</sup> gen	7 <sup>th</sup> gen
1991	1997	2000	2005	2008	2014	2020

© Ellis Horowitz & Marco Papa 1999-2024

Course Intro

51