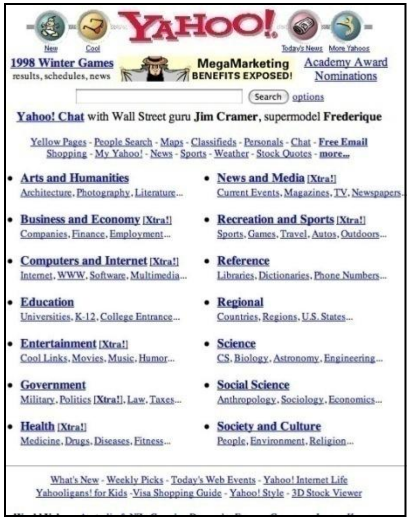
ICOM 6034
Website engineering

Dr. Roy Ho
Department of Computer Science, HKU

Session 10: Performance, traffic analysis and search engine optimizations

Scope

Part 1 (done):



New Cool YAHOO! Today's News More Yahoos

1998 Winter Games
results, schedules, news

MegaMarketing
BENEFITS EXPOSED!

Academy Award Nominations

Search options

Yahoo! Chat with Wall Street guru Jim Cramer, supermodel Frederique

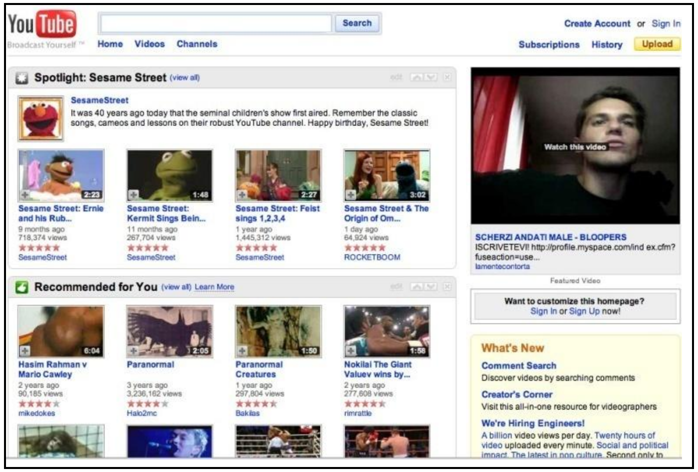
Yellow Pages - People Search - Maps - Classifieds - Personals - Chat - Free Email
Shopping - My Yahoo! - News - Sports - Weather - Stock Quotes - more...

- Arts and Humanities
  Architecture, Photography, Literature...
- Business and Economy [Xtra!]
  Companies, Finance, Employment...
- Computers and Internet [Xtra!]
  Internet, WWW, Software, Multimedia...
- Education
  Universities, K-12, College Entrance...
- Entertainment [Xtra!]
  Cool Links, Movies, Music, Humor...
- Government
  Military, Politics [Xtra!], Law, Taxes...
- Health [Xtra!]
  Medicine, Drugs, Diseases, Fitness...
- News and Media [Xtra!]
  Current Events, Magazines, TV, Newspapers...
- Recreation and Sports [Xtra!]
  Sports, Games, Travel, Autos, Outdoors...
- Reference
  Libraries, Dictionaries, Phone Numbers...
- Regional
  Countries, Regions, U.S. States...
- Science
  CS, Biology, Astronomy, Engineering...
- Social Science
  Anthropology, Sociology, Economics...
- Society and Culture
  People, Environment, Religion...

What's New - Weekly Picks - Today's Web Events - Yahoo! Internet Life
Yahooligans! for Kids -Visa Shopping Guide - Yahoo! Style - 3D Stock Viewer

Websites becoming more sophisticated, lots of interactivity

Part 2 (done):



Spotlight: Sesame Street (view all)

SesameStreet
It was 40 years ago today that the seminal children's show first aired. Remember the classic songs, cameos and lessons on their robust YouTube channel. Happy birthday, Sesame Street!

Sesame Street: Ernie and his Rub... 2:23
9 months ago
718,374 views
SesameStreet

Sesame Street: Kermit Sings Bein... 1:48
11 months ago
267,704 views
SesameStreet

Sesame Street: Feist sings 1,2,3,4 2:27
1 year ago
1,445,312 views
SesameStreet

Sesame Street & The Origin of Om... 3:02
1 day ago
64,924 views
ROCKETBOOM

Watch this video

SCHERZI ANDATI MALE - BLOOPERS
ISCRIVETEVI! http://profile.myspace.com/ind ex.cfm?fuseaction=use...
lamentecontorta

Featured Video

Want to customize this homepage?
Sign In or Sign Up now!

Recommended for You (view all) Learn More

Hasim Rahman v Mario Cawley 6:04
2 years ago
90,185 views
mikedokes

Paranormal 2:05
3 years ago
3,236,162 views
Halo2mc

Paranormal Creatures 1:50
1 year ago
297,804 views
Bakilas

Nokilai The Giant Valuev wins by... 1:58
2 years ago
277,608 views
rimrattle

What's New
Comment Search
Discover videos by searching comments
Creator's Corner
Visit this all-in-one resource for videographers
We're Hiring Engineers!
A billion video views per day. Twenty hours of video uploaded every minute. Social and political impact. The latest in pop culture. Second only to

Full-featured websites/apps are everywhere; integration and interoperability issues

Part 3 (done):

YouTube, Gmail,
Amazon, online
databases, Maps,
updated event lists,
YOUR websites,
....

The web/cloud(s)

Part 4 (Lecture 10):

You have a great website, how to make it loaded fast at users' computers, and most important… popular?

Optimizations

Session objectives

- Web page designs for faster rendering (i.e., better **performance**)
  - Goal: to minimize the rendering time (i.e., the loading time) of web pages in browsers
- Enhancing **popularity** of websites:
  - Web traffic analysis
    - Understand the users' behaviors and retain them
    - Introduction to Google Analytics
  - Search engine optimizations (SEO)
    - For improving the ranking of a website in search engines
- Course summary
- Exam format
- A short demo on Google Analytics

Web page designs for performance (faster page rendering)

(More in “Even Faster Web Sites” by Steve Souders)

Some techniques for making web pages load faster

1. Make fewer HTTP requests
2. Use a CDN
3. Add an Expires header
4. Gzip components
5. Put CSS at the top
6. Move JS to the bottom
7. Make JS and CSS external
8. Reduce DNS lookups
9. Minify JS
10. Avoid redirects

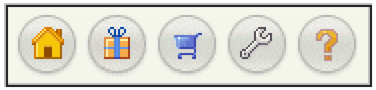
1: Make fewer HTTP requests

- Latency (compared to bandwidth) is the dominating factor of client-perceived performance
  - Consider the bandwidth and the client-perceived latency in mobile (e.g., 4G) networks
  - 5G better, but far less common, and is prone to coverage/reception problems as well

- **And**: each HTTP request = a round-trip latency
  - => avoiding HTTP requests is an effective way to improve client-perceived performance

- Main techniques:
  - Use image maps instead of multiple image files
  - Combine scripts and stylesheets

Image maps



- Client-side
  - HTML:

    ```
    <img usemap="#map1" border=0 src="/images/imagemap.gif">
    <map name="map1">
      <area shape="rect" coords="0,0,31,31" href="home.html" title="Home">
        …
    </map>
    ```

  - Or use an appropriate plugin for your chosen client-side library/framework (e.g., jQuery)
- Server-side
  - `<a href="navbar.php"><img ismap src="imagemap.gif"></a>`
  - When the above link is clicked, the coordinates of the click (inside the image map) would be appended to the query string
  - → `http://.../navbar.php?127,13`
  - `(but the client-side approach is more commonly used)`

Combine scripts / stylesheets

	Scripts	Stylesheets
amazon.com	3	1
aol.com	18	1
cnn.com	11	2
ebay.com	7	2
msn.com	9	1
myspace.com	2	2
wikipedia.org	3	1
yahoo.com	4	1
youtube.com	7	3
Average	**6.5**	**1.5**

Combined scripts / stylesheets

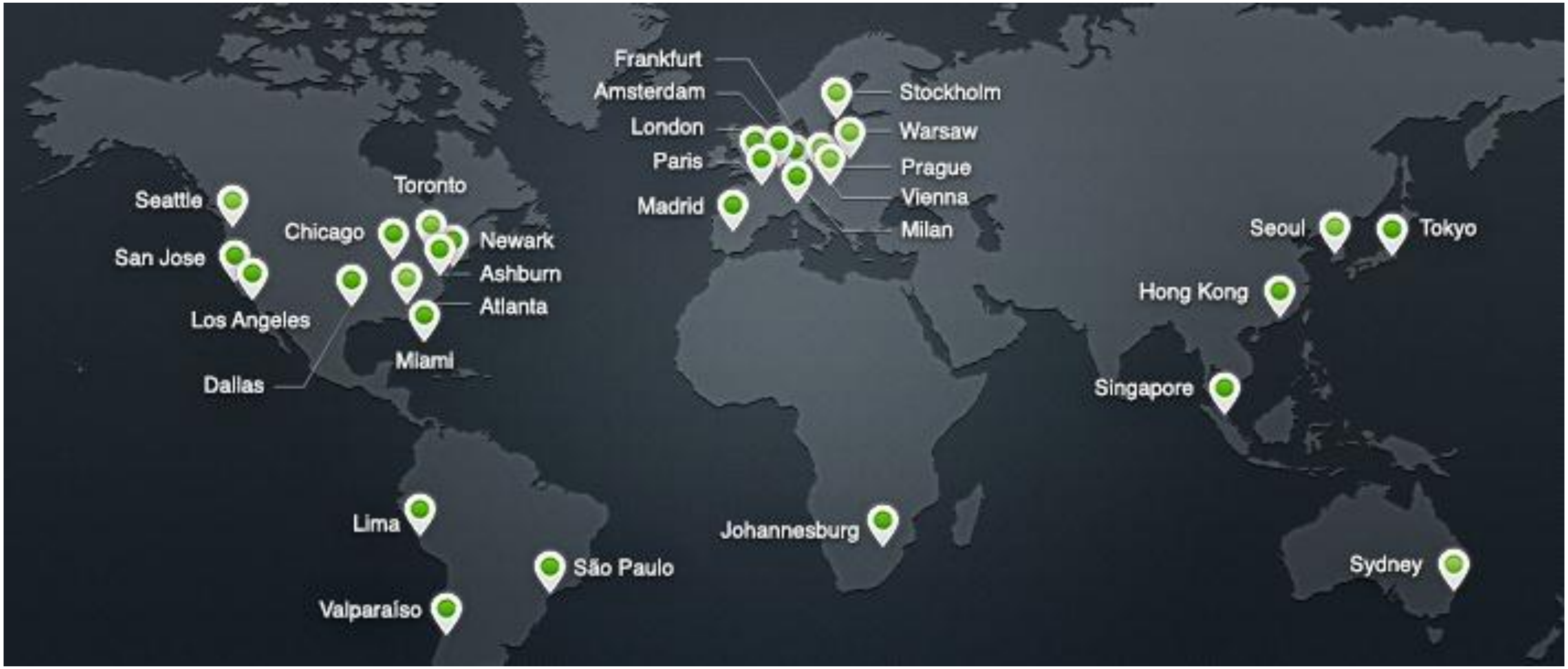
- Combining six scripts into one eliminates five HTTP requests (i.e., five round-trip latencies)
- Challenge: scripts are often developed as separate modules by different teams/developers
- One solution:
  - Dynamically combine multiple scripts into one right before delivery

2: Use a Content Delivery Network (CDN)

- A CDN (e.g., Akamai, CloudFlare, etc.) helps cache a website's content in geographically-distributed server networks
  - Many cloud providers (e.g., Google/Amazon) and CDNs have regional data centers
  - So the contents can be placed "closer" to clients, while the clients can obtain the contents with less network delays
- If your website is popular and have international audiences, using CDNs should be *very* beneficial for improving the user experience

amazon.com	Akamai
aol.com	Akamai
cnn.com	
ebay.com	Akamai, Mirror Image
google.com	
msn.com	SAVVIS
myspace.com	Akamai, Limelight
wikipedia.org	
yahoo.com	Akamai
youtube.com	

An example of CDN: CloudFlare



- Geographically-distributed CDN servers as shown above
- CloudFlare claims that it helps websites to handle 65% of their requests on average, with client-perceived latency being cut to half
- Free-of-charge for basic set-up; gaining popularity in the recent years
- Additional services are subscription-based. E.g., protection from DDoS attack costs ~US$200/month for each website.

3: Add an “expires” HTTP header

- An “expires” HTTP header tells the browser’s or proxy’s cache to keep the files (and reuse them) before they are expired => avoid unnecessary retrievals
  - Can be specified with the header() function in PHP
- Responses to AJAX can also be cached.
- Components that are set cacheable for 30 days or more in large websites:

	Image	**Stylesheet**	**Script**	%	Median Age
aol.com	23/43	1/1	6/18	48%	217 days
cnn.com	2/138	0/2	2/11	1%	227 days
ebay.com	16/20	0/2	0/7	55%	140 days
msn.com	32/35	1/1	3/9	80%	34 days
myspace.com	0/18	0/2	0/2	0%	1 day
wikipedia.org	6/8	1/1	2/3	75%	1 day
yahoo.com	23/23	1/1	4/4	100%	60 days
youtube.com	2/32	0/3	0/7	0%	26 days

4: Gzip components

- Compress the data before sending to clients; client browsers decompress them before rendering

- All popular browsers support decompression

- Configuration at the server side:
  - Apache: mod_deflate

    ```
    AddOutputFilterByType DEFLATE text/html text/css application/x-javascript
    ```

  - HTTP response

    ```
    Content-Encoding: gzip
    ```

Gzip: not just for HTML

	HTML	Scripts	Stylesheets
amazon.com	x		
aol.com	x	some	Some
cnn.com			
ebay.com	x		
msn.com	x	some	some
myspace.com	x	x	x
wikipedia.org	x	x	x
yahoo.com	x	x	x
youtube.com	x	some	some

- gzip scripts, stylesheets, XML, JSON
- Images, PDF, etc. are often not compressed

5: Put CSS at the top of a page

- Put CSS in html's HEAD section using the LINK tag

- => HTML elements (e.g., an image) can be placed and styled properly once they arrive
  - => users would **feel** faster rendering
  - Can avoid flashing of unstyled contents as that arrive

  - (But actually the total loading time for all elements is the same)

6: Move JS to the bottom

- Many JS manipulate the DOM tree (i.e., add/remove DOM elements)
  - => Once a script is encountered by the browser, the rendering of the page would be paused until all components (e.g., images) arrive and the script's execution is finished
  - That's why some pages appear to be "blank" when loading (which can take long), but suddenly the entire page with all components appears
  - => Long client-perceived latency
- Scripts also block parallel downloads across all hostnames until they finish their execution
- Solution: move JS as low in the page as possible

7: Make JS and CSS external

- JS/CSS are far less likely to be modified compared to HTML

- Make JS/CSS external => more HTTP requests, but **cacheable**
  - The HTML will also become smaller

- Further technique: Use Post-Onload download to prefetch some essential JS/CSS files for future use (e.g., when the users navigate “deeper” in the site)
  - download external files after onload

```
window.onload = downloadComponents;
function downloadComponents() {
    var elem = document.createElement("script");
    elem.src = "http://.../file1.js";
    document.body.appendChild(elem);
    ...
}
```

  - => can speed up the display of “secondary-level” pages!

8: Reduce DNS lookups

- A DNS lookup typically takes 20-120 ms
- Use fewer hostnames – 2-4
  - E.g., hostname[1-4].yoursite.com
- Use the HTTP header: keep-alive
  - Multiple HTTP requests can be sent in one TCP/HTTP connection
  - The total number of new connections (which may result in DNS lookups) can be significantly reduced
  - Enabled by default in all newer browsers and web servers

9: Minify JavaScript and CSS

- Raw JS/CSS is wordy, transfer consumes network bandwidth
- Minify: remove all unnecessary characters

	Minify External?	**Minify Inline?**
www.amazon.com	no	no
www.aol.com	no	no
www.cnn.com	no	no
www.ebay.com	yes	no
www.msn.com	yes	yes
www.myspace.com	no	no
www.wikipedia.org	no	no
www.yahoo.com	yes	yes
www.youtube.com	no	no

- minify inline scripts also help

Minify

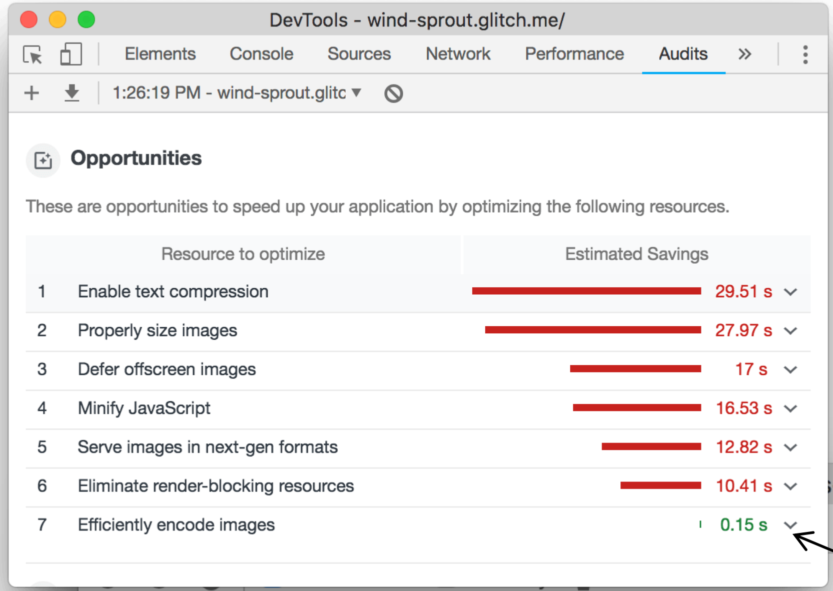
	Original	JSMin savings	Dojo savings
www.amazon.com	204K	31K (15%)	48K (24%)
www.aol.com	44K	4K (10%)	4K (10%)
www.cnn.com	98K	19K (20%)	24K (25%)
www.myspace.com	88K	23K (27%)	24K (28%)
www.wikipedia.org	42K	14K (34%)	16K (38%)
www.youtube.com	34K	8K (22%)	10K (29%)
Average	**85K**	**17K (21%)**	**21K (25%)**

- Example tool:
  - https://www.minifier.org/ (which minifies JS & CSS)

10. Avoid redirections

- HTTP redirections take double round-trip time
- E.g., a HTTP redirection would occur when a trailing slash (/) is missing from a URL that should otherwise have one, e.g.,
- The URL: http://example.com/folder would result in a 301 response (a redirection) if it should contain a slash (/): http://example.com/folder/
  - => One of the most common (and avoidable) redirections, but developers are generally not aware of it.

Chrome DevTools and Firefox Developer Tools

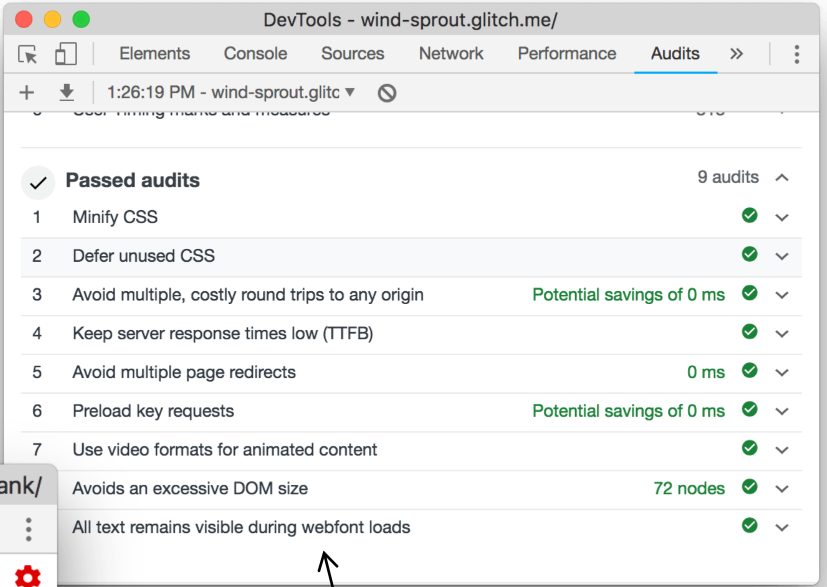


DevTools - wind-sprout.glitch.me/

Opportunities

These are opportunities to speed up your application by optimizing the following resources.

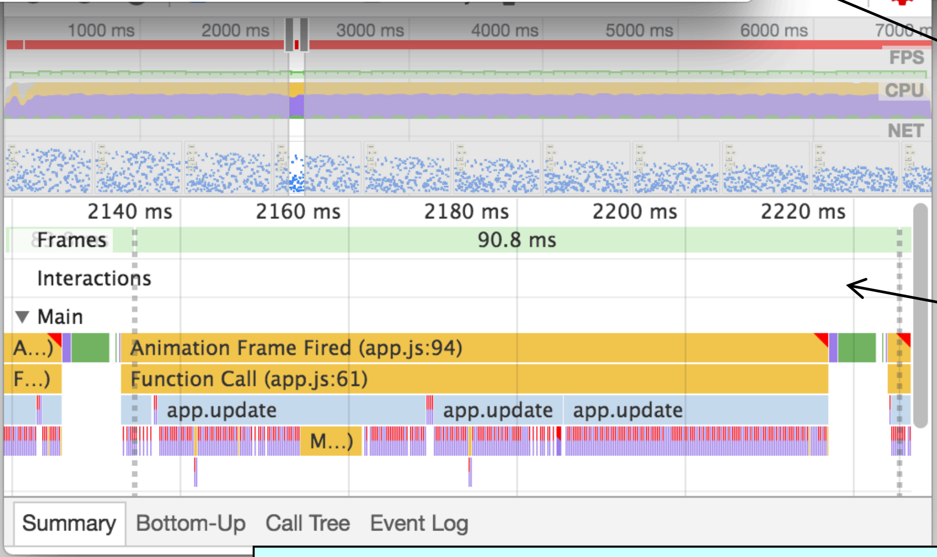
	Resource to optimize	Estimated Savings
1	Enable text compression	29.51 s
2	Properly size images	27.97 s
3	Defer offscreen images	17 s
4	Minify JavaScript	16.53 s
5	Serve images in next-gen formats	12.82 s
6	Eliminate render-blocking resources	10.41 s
7	Efficiently encode images	0.15 s



DevTools - wind-sprout.glitch.me/

Passed audits 9 audits

1	Minify CSS		✓
2	Defer unused CSS		✓
3	Avoid multiple, costly round trips to any origin	Potential savings of 0 ms	✓
4	Keep server response times low (TTFB)		✓
5	Avoid multiple page redirects	0 ms	✓
6	Preload key requests	Potential savings of 0 ms	✓
7	Use video formats for animated content		✓
	Avoids an excessive DOM size	72 nodes	✓
	All text remains visible during webfont loads		✓



Analyze your website (e.g., its loading and running times) and suggest ways to optimize it

Detailed profiling provide insights into your website's performance, e.g., how much time is spent in each event / function call, etc.

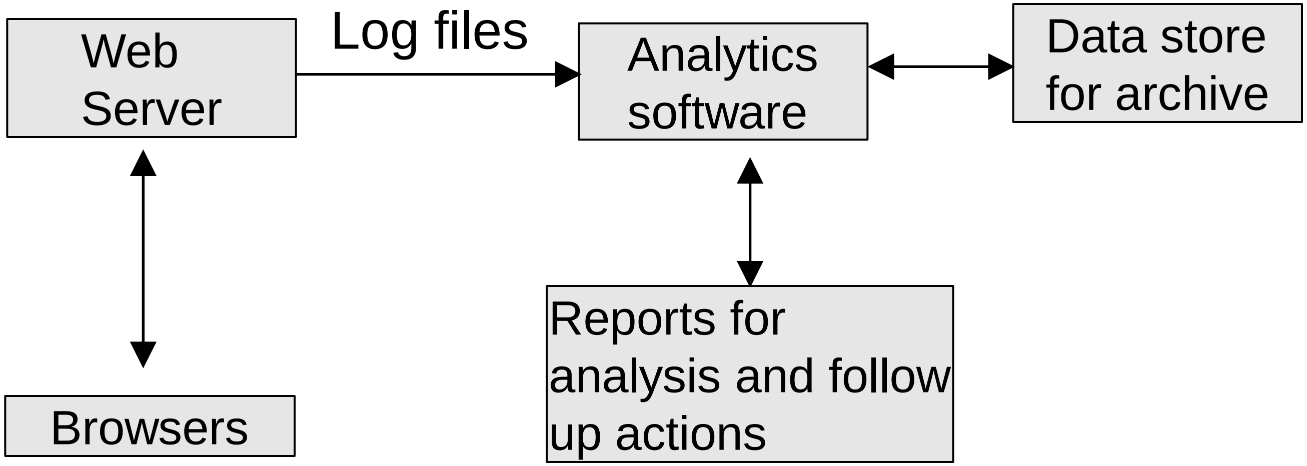
Chrome: https://developers.google.com/web/tools/chrome-devtools/evaluate-performance/
Firefox: https://developer.mozilla.org/en-US/docs/Tools

Traffic analysis

Definition

Web traffic analysis is the process of understanding the behaviours of clients on a website, and applying the conclusions and follow-up actions.

Goal: to optimize the user experiences, retain them and achieve the purpose of the website (e.g., sales of products)



Two main techniques

1. Use a log file analysis software
   - □ "Data-mine" a web server's log files to analyze/understand the traffic

2. Apply page tagging with a third-party, hosted, analytics service
   - □ Include a link to an external (but invisible) object in every (or some important) web pages.
   - □ That external object is hosted in an analytics service provider
   - □ When your page is accessed by a client, that external object would be accessed as well
   - □ The service provider counts how many times this object is retrieved, and when it is retrieved.
     - ■ Tools are available for analyzing and visualizing the traffic
   - □ => a "cloud version" of the traditional logging approach

- ■ List of web analytics software:
  - □ http://en.wikipedia.org/wiki/List_of_web_analytics_software

Log file example
(Combined Log Format)

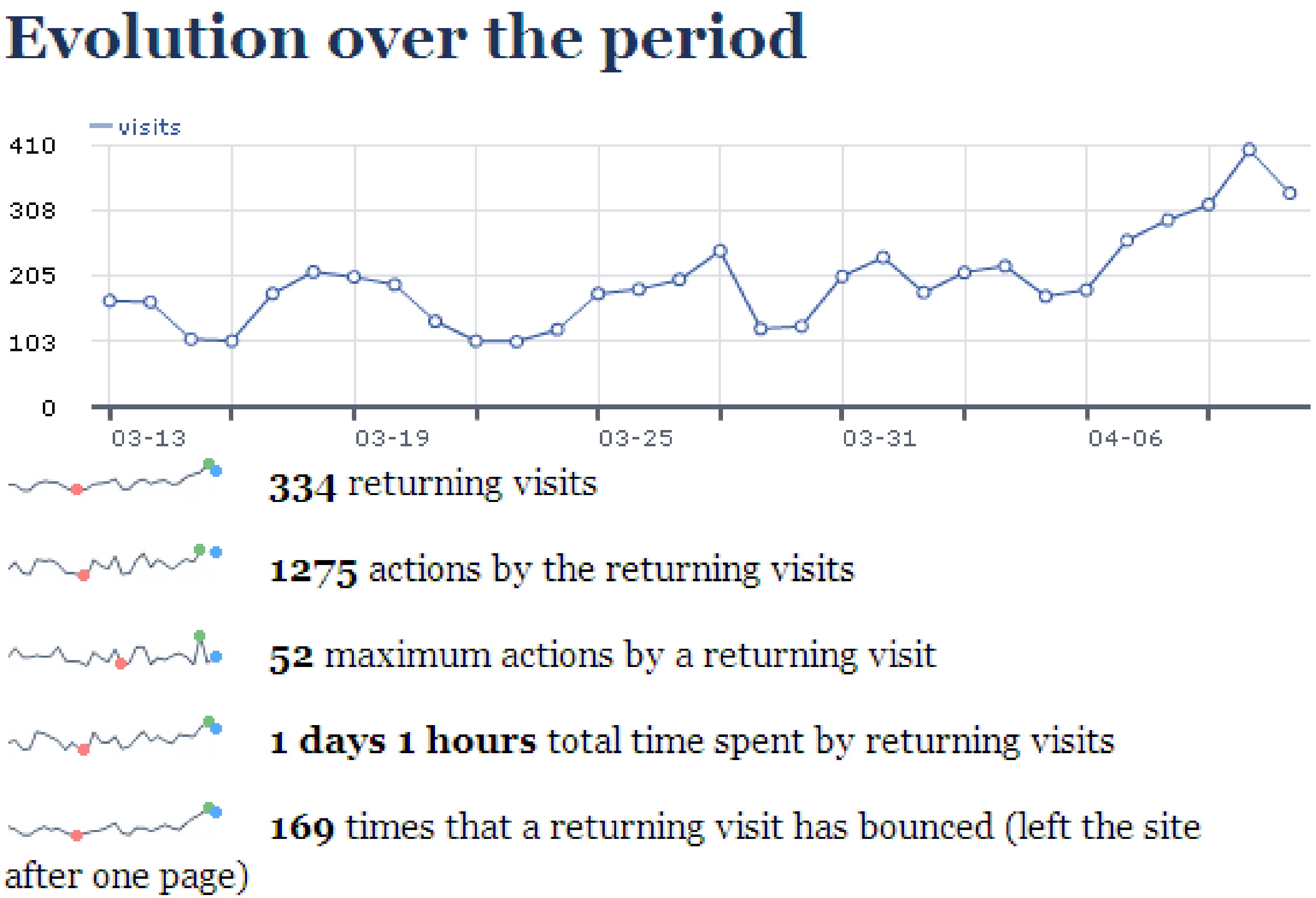
```
ecomicom.hku.hk user - [19/Jun/2018:00:49:41 - 0500] "GET
/service/contracts.jpg HTTP/1.1" 200 1341 "https://www.google.com/"
"Firefox/55 (Windows 10)"
```

- Hostname or IP of client
- Registered user name (usually blank or “`nobody`”)
- Date and time of request
- Object requested
- HTTP status code
- Bytes transferred
- Referral information
- Browser/platform information

Some opensource analysis tools: Piwik (now called “Matomo”), AWStats

More: https://www.cyberciti.biz/open-source/7-awesome-open-source-analytics-weblog-analysis-softwares/

Examples of report: request counts



Evolution over the period

334 returning visits

1275 actions by the returning visits

52 maximum actions by a returning visit

1 days 1 hours total time spent by returning visits

169 times that a returning visit has bounced (left the site after one page)

Top pages report

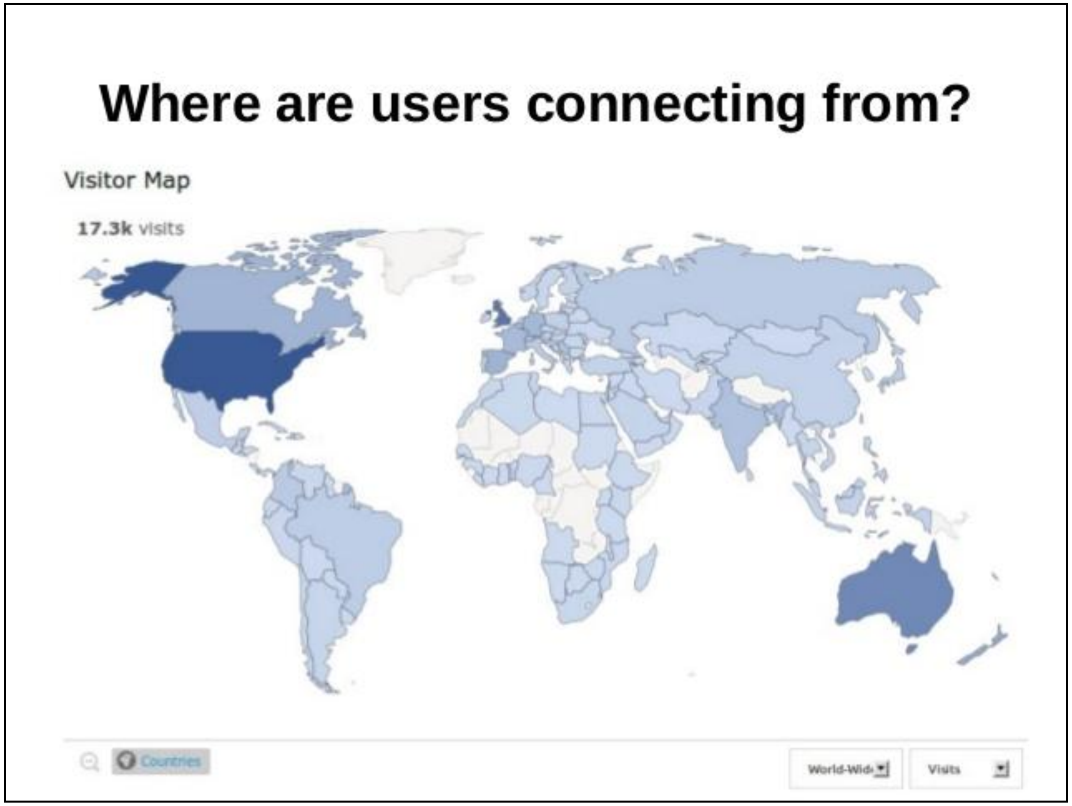
Important for understanding which pages in the website attract, and which do not

Title Never Resolved
Title Never Resolved
Title Never Resolved
Title Never Resolved
Title Never Resolved
Title Never Resolved
Title Never Resolved
Title Never Resolved
Title Never Resolved
<= 1.0%

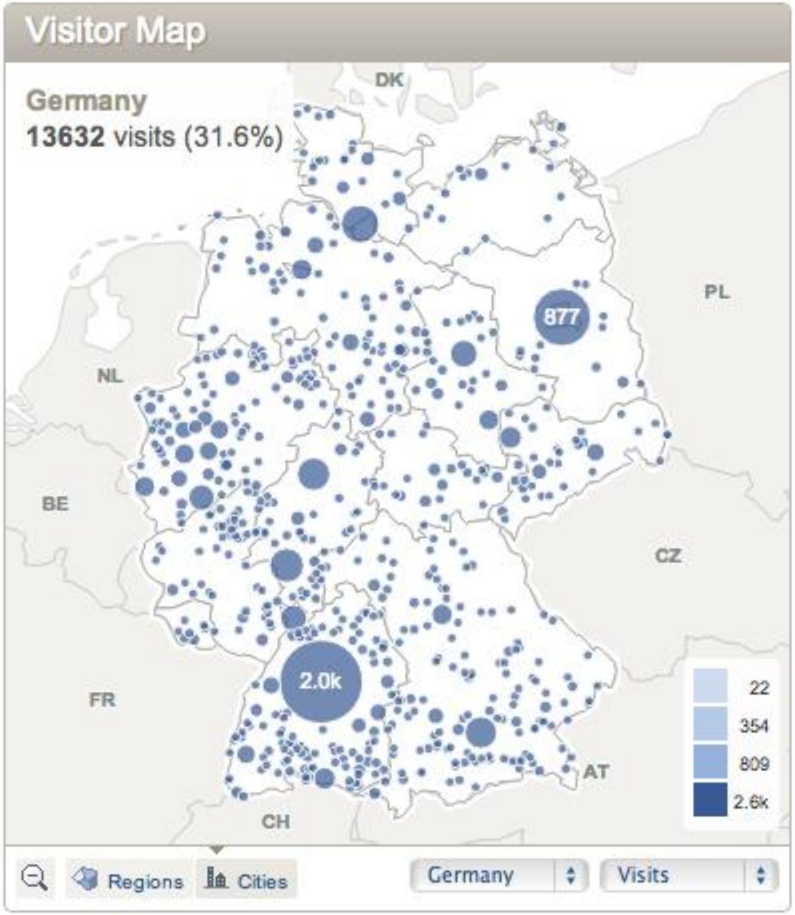
HTTP Resource	# of Page Views	%of Total	Cum %
1. /	2,136,650	38.5	38.5
2. /bigadmin/downloads/	228,679	4.1	42.7
3. /MySun/	198,430	3.6	46.2
4. /bigadmin/docs/	131,694	2.4	48.6
5. /search/index.php	103,248	1.9	50.5
6. /openoffice/	65,347	1.2	51.6
7. /products-n-solutions	65,038	1.2	52.8
8. /corp_emp/scripts/openings.php	63,601	1.1	54.0
9. /products/openoffice/get.php	60,260	1.1	55.0.
10. /forte/ffj/overview.html/	58,103	1.0	56.1
11. Others	2,434,788	43.9	100

Top regions

Important for understanding the physical location (country or even organization) of clients

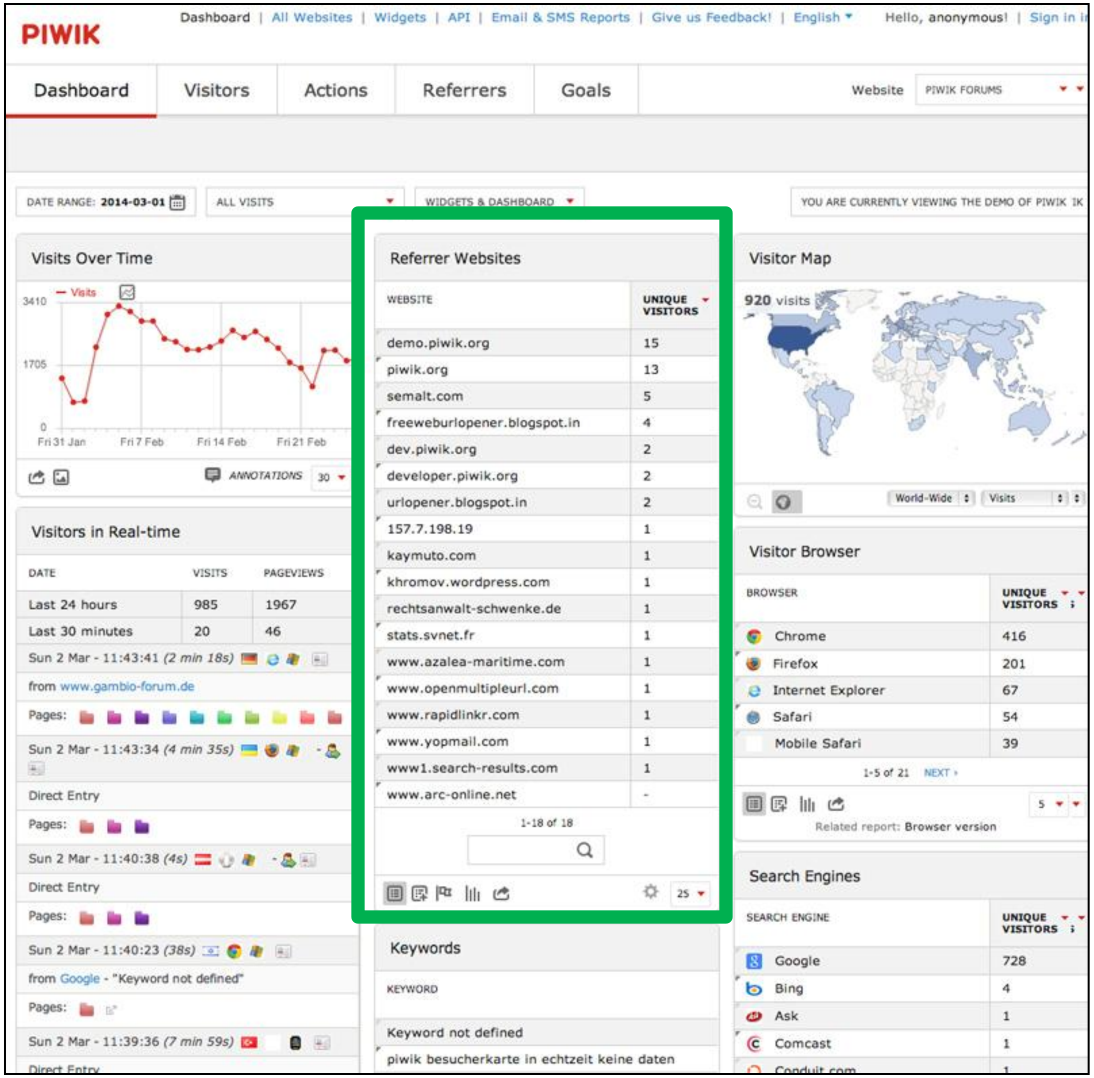


Where are users connecting from?



Top referrers

Important for analyzing results of web marketing campaigns



PIWIK

Dashboard | All Websites | Widgets | API | Email & SMS Reports | Give us Feedback! | English ▾ Hello, anonymous! | Sign in

Dashboard | Visitors | Actions | Referrers | Goals — Website: PIWIK FORUMS

DATE RANGE: 2014-03-01 | ALL VISITS | WIDGETS & DASHBOARD | YOU ARE CURRENTLY VIEWING THE DEMO OF PIWIK

Visits Over Time

Visitors in Real-time

DATE	VISITS	PAGEVIEWS
Last 24 hours	985	1967
Last 30 minutes	20	46

Sun 2 Mar - 11:43:41 (2 min 18s)

from www.gambio-forum.de

Pages:

Sun 2 Mar - 11:43:34 (4 min 35s)

Direct Entry

Pages:

Sun 2 Mar - 11:40:38 (4s)

Direct Entry

Pages:

Sun 2 Mar - 11:40:23 (38s)

from Google - "Keyword not defined"

Pages:

Sun 2 Mar - 11:39:36 (7 min 59s)

Direct Entry

Referrer Websites

WEBSITE	UNIQUE VISITORS
demo.piwik.org	15
piwik.org	13
semalt.com	5
freeweburlopener.blogspot.in	4
dev.piwik.org	2
developer.piwik.org	2
urlopener.blogspot.in	2
157.7.198.19	1
kaymuto.com	1
khromov.wordpress.com	1
rechtsanwalt-schwenke.de	1
stats.svnet.fr	1
www.azalea-maritime.com	1
www.openmultipleurl.com	1
www.rapidlinkr.com	1
www.yopmail.com	1
www1.search-results.com	1
www.arc-online.net	-

1-18 of 18

25

Keywords

KEYWORD
Keyword not defined
piwik besucherkarte in echtzeit keine daten

Visitor Map

920 visits

World-Wide | Visits

Visitor Browser

BROWSER	UNIQUE VISITORS
Chrome	416
Firefox	201
Internet Explorer	67
Safari	54
Mobile Safari	39

1-5 of 21 NEXT ›

5

Related report: Browser version

Search Engines

SEARCH ENGINE	UNIQUE VISITORS
Google	728
Bing	4
Ask	1
Comcast	1
Conduit.com	1

Clickstream

Important for understanding the users' navigation, identifying the "drop-out" points, and improving website design to retain clients – similar to "conversion funnels" (more later)

	Number of Visits			% of Tota
First page	11507			100.00%
Second page		9096		79.00%
Third page			7000	61.00%
Third page			1500	13.00%
Third page			500	4.30%
Third page			96	0.80%
Second page		1214		10.60%
Third page			577	5.00%
Third page			394	3.40%
Third page			134	1.20%
Third page			109	0.90%
Second page		1137		9.90%
Third page			800	7.00%
Third page			200	1.70%
Third page			100	0.90%
Third page			37	0.30%
Second page		20		0.20%
Third page			10	0.10%
Third page			5	0.05%
Third page			3	0.03%

Some issues to consider

- Browser and proxy caching
  - Affect total quantities of traffic (views, users)
  - Affect apparent behavior (e.g., click streams)
  - Many real clients behind a proxy server may appear as one client
    - Solution: embed an unique ID for each user in the URL or cookies, and configure the analytics software to recognize user sessions

- Multiple servers for a website
  - E.g., a cluster-based web server (or called a “web server farm”)
  - Need to merge all log files stored in multiple server nodes.
  - Some advanced analytics software (e.g., Piwik/Matomo) can support such cluster-based environments

- Robots
  - Page view counts increased unintentionally
  - Solution: set the analytics software to ignore or treat robots’ requests in a special way

Some issues to consider

Dynamic content:

The URLs of dynamic pages can be very cryptic (e.g., with long query strings).

Most analytics software support dynamic patterns of URLs, but need to be set properly

Query Strings:

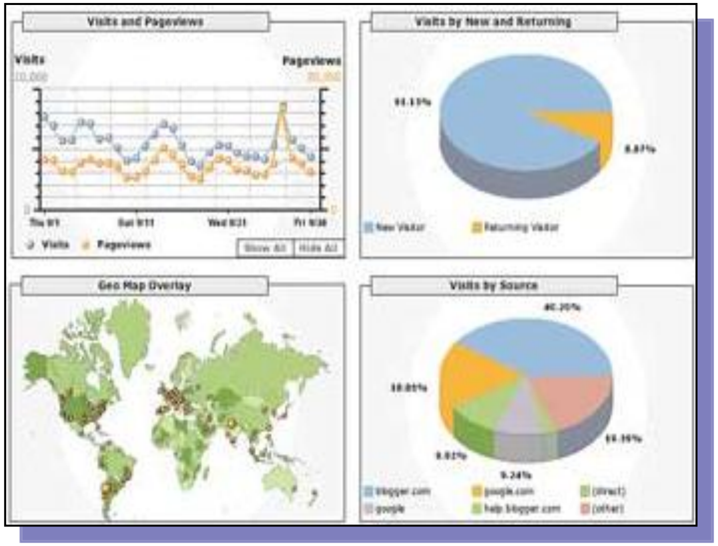
RESOURCE?KEY=*value*&KEY=*value*
https://oracle.com/service/Router?country=US
https://oracle.com/service/Router?country=JP

Conclusion: *one must know the behaviour of your analytics software on these aspects,* ***and configure it properly****, in order to have correct interpretation of the analysis results*.

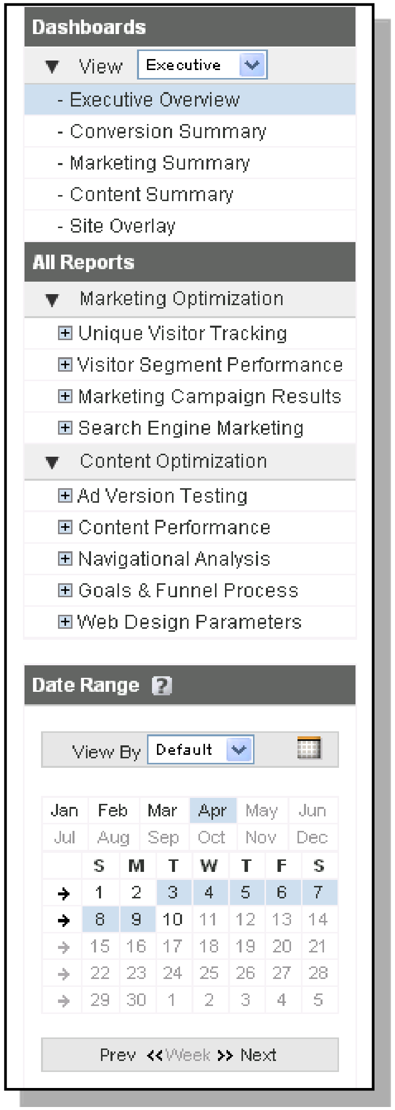
Example of hosted analytics services: Google Analytics

A powerful analysis tool

- A page tagging technique for web traffic analysis
- Free of charge
- Over 80 types of reports
- Main advantages:
  - Measure how many clients you have and their behaviour
  - Measure e-commerce transactions
  - Measure effects of online marketing campaigns
  - All reports available online



Dashboards Layout and Navigation



Dashboards

▼ View Executive

- Executive Overview
- Conversion Summary
- Marketing Summary
- Content Summary
- Site Overlay

All Reports

▼ Marketing Optimization

- ⊞ Unique Visitor Tracking
- ⊞ Visitor Segment Performance
- ⊞ Marketing Campaign Results
- ⊞ Search Engine Marketing

▼ Content Optimization

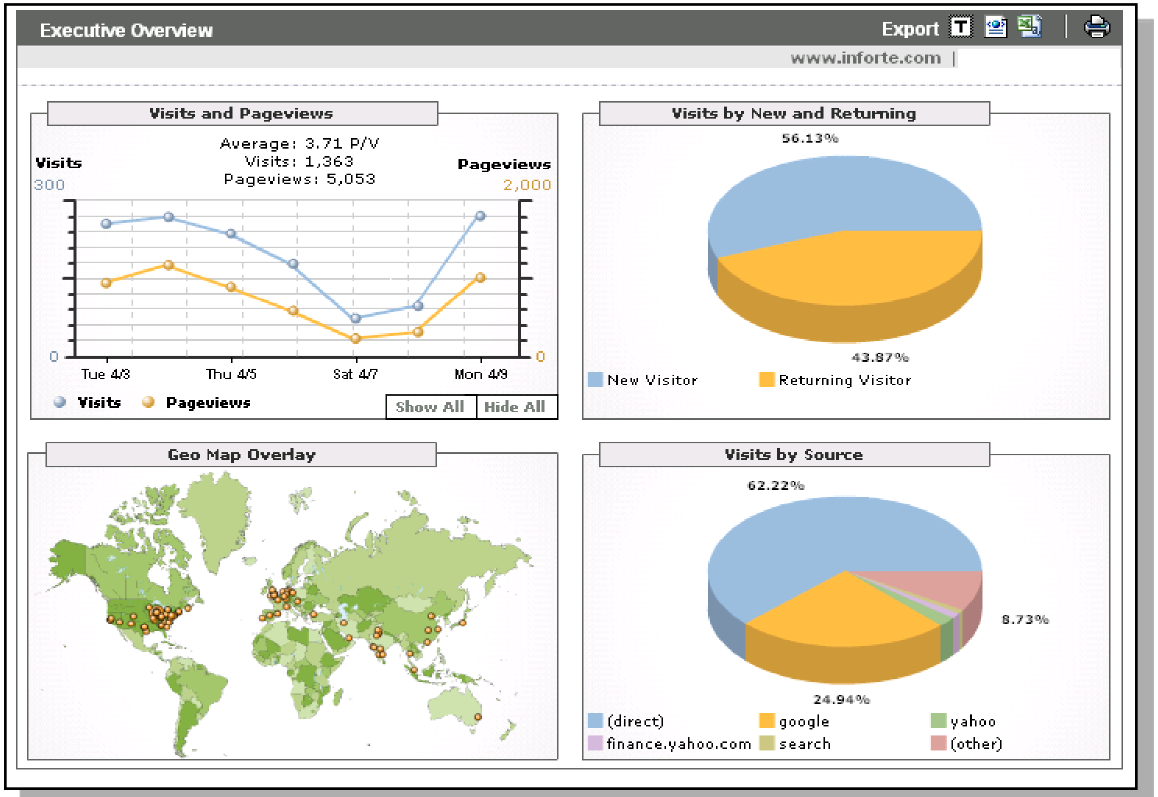
- ⊞ Ad Version Testing
- ⊞ Content Performance
- ⊞ Navigational Analysis
- ⊞ Goals & Funnel Process
- ⊞ Web Design Parameters

Date Range ?

View By Default

Jan	Feb	Mar	Apr	May	Jun	
Jul	Aug	Sep	Oct	Nov	Dec	
	S	M	T	W	T	F	S
→	1	2	3	4	5	6	7
→	8	9	10	11	12	13	14
→	15	16	17	18	19	20	21
→	22	23	24	25	26	27	28
→	29	30	1	2	3	4	5

Prev << Week >> Next



Executive Overview Export

www.inforte.com |

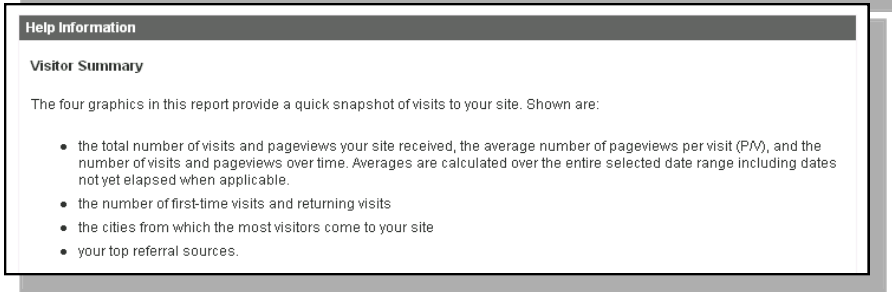
Visits and Pageviews

Average: 3.71 P/V
Visits: 1,363
Pageviews: 5,053

Visits by New and Returning

Geo Map Overlay

Visits by Source



Help Information

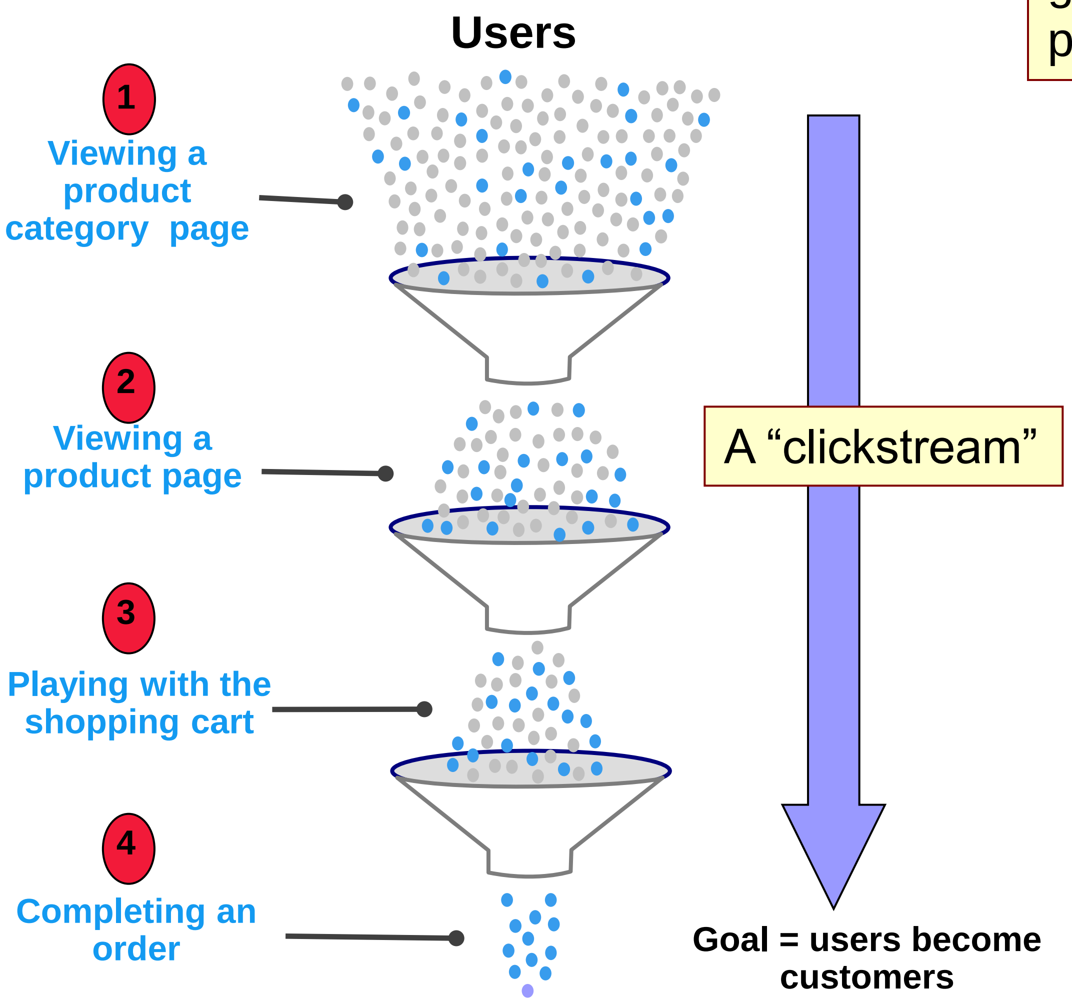
Visitor Summary

The four graphics in this report provide a quick snapshot of visits to your site. Shown are:

- the total number of visits and pageviews your site received, the average number of pageviews per visit (P/V), and the number of visits and pageviews over time. Averages are calculated over the entire selected date range including dates not yet elapsed when applicable.
- the number of first-time visits and returning visits
- the cities from which the most visitors come to your site
- your top referral sources.

Conversion funnel

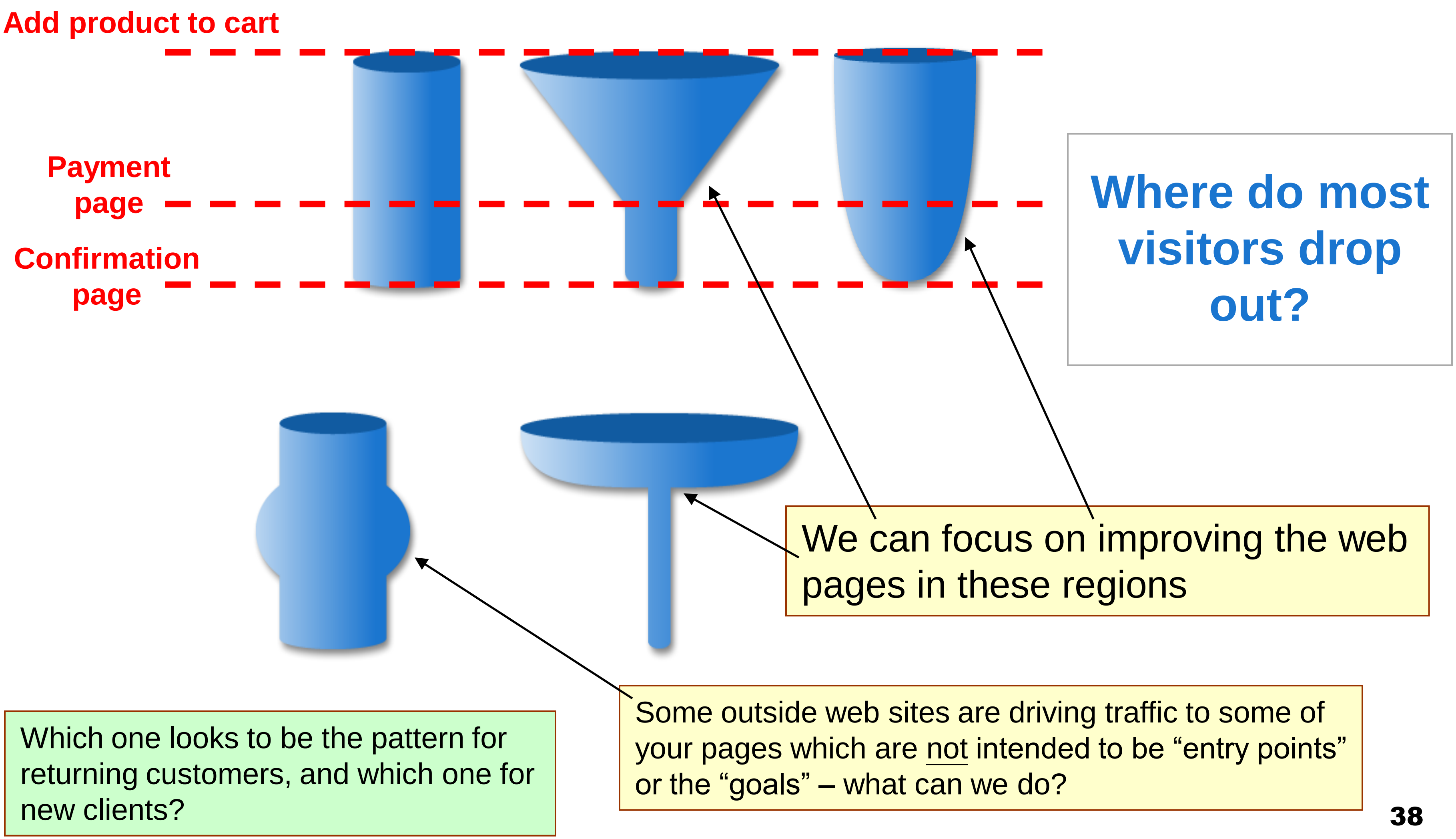
“Conversion” – the process of converting an access (e.g., to your front page) to an action that achieves the goal of the website (e.g., placing an order).


Users
1
Viewing a product category page
2
Viewing a product page
3
Playing with the shopping cart
4
Completing an order
A “clickstream”
Goal = users become customers

Where do most visitors drop out?

Optimize conversion process to maximize conversions

Some examples of drop-out rates



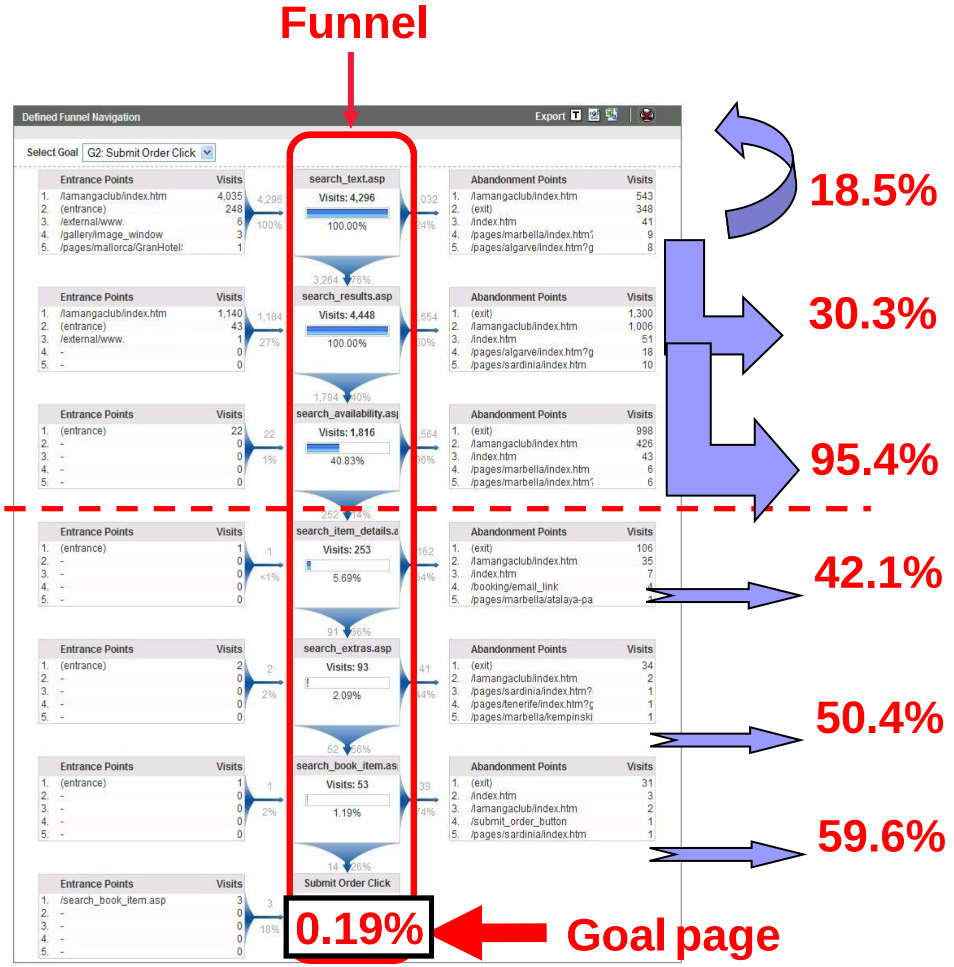
Where do most visitors drop out?

We can focus on improving the web pages in these regions

Some outside web sites are driving traffic to some of your pages which are not intended to be “entry points” or the “goals” – what can we do?

Which one looks to be the pattern for returning customers, and which one for new clients?

Example: conversion funnel of a flight reservation service



Funnel

Defined Funnel Navigation

Export

Select Goal: G2: Submit Order Click

Search page 1

	Entrance Points	Visits
1.	/lamangaclub/index.htm	4,035
2.	(entrance)	248
3.	/external/www.	6
4.	/gallery/image_window	3
5.	/pages/mallorca/GranHotel:	1

4,296 100%

search_text.asp

Visits: 4,296

100.00%

,032 24%

	Abandonment Points	Visits
1.	/lamangaclub/index.htm	543
2.	(exit)	348
3.	/index.htm	41
4.	/pages/marbella/index.htm?	9
5.	/pages/algarve/index.htm?g	8

18.5%

3,264 76%

Results page 2

	Entrance Points	Visits
1.	/lamangaclub/index.htm	1,140
2.	(entrance)	43
3.	/external/www.	1
4.	-	0
5.	-	0

1,184 27%

search_results.asp

Visits: 4,448

100.00%

,654 60%

	Abandonment Points	Visits
1.	(exit)	1,300
2.	/lamangaclub/index.htm	1,006
3.	/index.htm	51
4.	/pages/algarve/index.htm?g	18
5.	/pages/sardinia/index.htm	10

30.3%

1,794 40%

Flight availability 3

	Entrance Points	Visits
1.	(entrance)	22
2.	-	0
3.	-	0
4.	-	0
5.	-	0

22 1%

search_availability.as|

Visits: 1,816

40.83%

,564 86%

	Abandonment Points	Visits
1.	(exit)	998
2.	/lamangaclub/index.htm	426
3.	/index.htm	43
4.	/pages/marbella/index.htm	6
5.	/pages/marbella/index.htm?	6

95.4%

252 14%

Reservation 4

	Entrance Points	Visits
1.	(entrance)	1
2.	-	0
3.	-	0
4.	-	0
5.	-	0

1 <1%

search_item_details.a

Visits: 253

5.69%

162 64%

	Abandonment Points	Visits
1.	(exit)	106
2.	/lamangaclub/index.htm	35
3.	/index.htm	7
4.	/booking/email_link	4
5.	/pages/marbella/atalaya-pa	1

42.1%

91 36%

Confirmation 5

	Entrance Points	Visits
1.	(entrance)	2
2.	-	0
3.	-	0
4.	-	0
5.	-	0

2 2%

search_extras.asp

Visits: 93

2.09%

41 44%

	Abandonment Points	Visits
1.	(exit)	34
2.	/lamangaclub/index.htm	2
3.	/pages/sardinia/index.htm?	1
4.	/pages/tenerife/index.htm?g	1
5.	/pages/marbella/kempinski	1

50.4%

52 56%

Payment form 6

	Entrance Points	Visits
1.	(entrance)	1
2.	-	0
3.	-	0
4.	-	0
5.	-	0

1 2%

search_book_item.as|

Visits: 53

1.19%

39 74%

	Abandonment Points	Visits
1.	(exit)	31
2.	/index.htm	3
3.	/lamangaclub/index.htm	2
4.	/submit_order_button	1
5.	/pages/sardinia/index.htm	1

59.6%

14 26%

Order complete 7

	Entrance Points	Visits
1.	/search_book_item.asp	3
2.	-	0
3.	-	0
4.	-	0
5.	-	0

3 18%

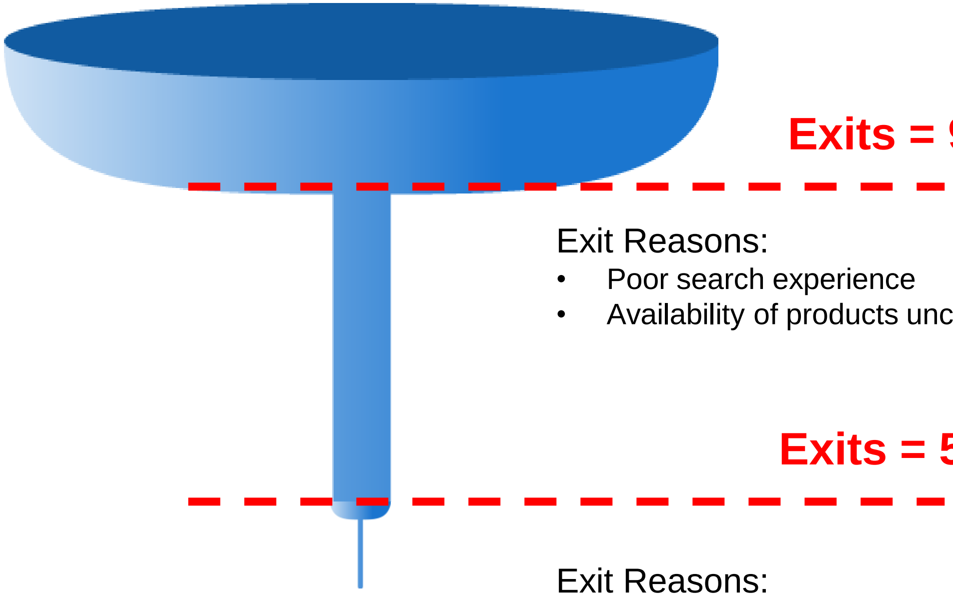
Submit Order Click

0.19%

Goal page

Where do most users drop out?

Analyzing the reasons for drop-out



Users

Exits = 95.4%

Exit Reasons:

- Poor search experience
- Availability of products unclear, etc.

Exits = 59.6%

Exit Reasons:

- Over-complicated payment form
- Poor layout, etc.

Customers

Identify reasons for high drop-out

The “exit reasons” can be found based on on-line surveys, or focus group discussions

Some reasons are trivial – e.g., inconvenient search/payment forms, slow searching, over-crowded search results, insufficient info for commitment, etc.

Improving the websites

1
Search page

2
Results page

3
Flight availability

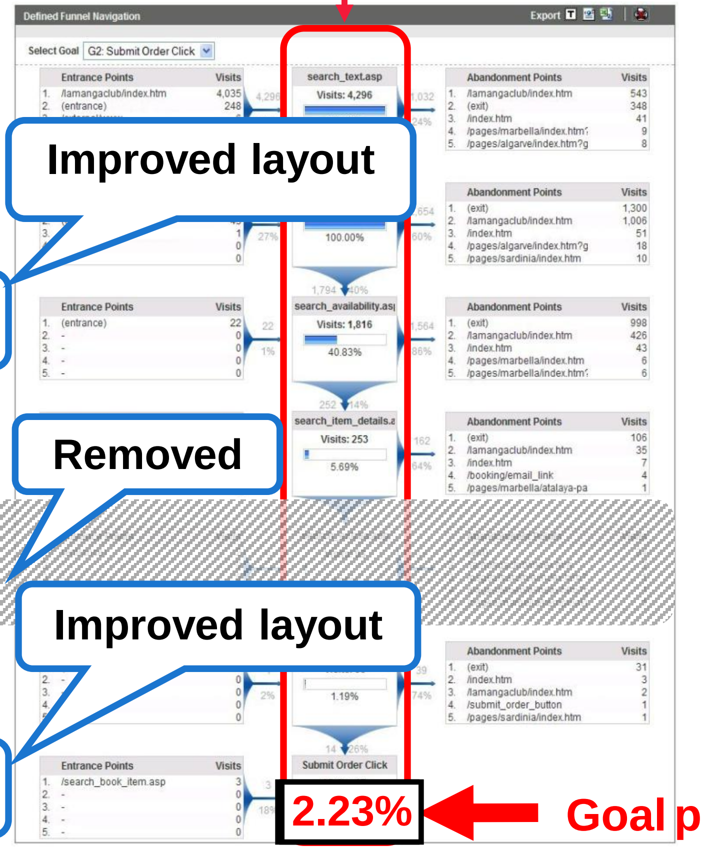
4
Reservation

5
Confirmation

6
Payment form

7
Order complete

Funnel



Defined Funnel Navigation

Export

Select Goal: G2: Submit Order Click

	Entrance Points	Visits
1.	/lamangaclub/index.htm	4,035
2.	(entrance)	248

search_text.asp
Visits: 4,296

	Abandonment Points	Visits
1.	/lamangaclub/index.htm	543
2.	(exit)	348
3.	/index.htm	41
4.	/pages/marbella/index.htm?	9
5.	/pages/algarve/index.htm?g	8

100.00%

	Abandonment Points	Visits
1.	(exit)	1,300
2.	/lamangaclub/index.htm	1,006
3.	/index.htm	51
4.	/pages/algarve/index.htm?g	18
5.	/pages/sardinia/index.htm	10

1,794 40%

	Entrance Points	Visits
1.	(entrance)	22
2.	-	0
3.	-	0
4.	-	0
5.	-	0

search_availability.asp
Visits: 1,816
40.83%

	Abandonment Points	Visits
1.	(exit)	998
2.	/lamangaclub/index.htm	426
3.	/index.htm	43
4.	/pages/marbella/index.htm	6
5.	/pages/marbella/index.htm?	6

252 14%

search_item_details.a
Visits: 253
5.69%

	Abandonment Points	Visits
1.	(exit)	106
2.	/lamangaclub/index.htm	35
3.	/index.htm	7
4.	/booking/email_link	4
5.	/pages/marbella/atalaya-pa	1

1.19%

	Abandonment Points	Visits
1.	(exit)	31
2.	/index.htm	3
3.	/lamangaclub/index.htm	2
4.	/submit_order_button	1
5.	/pages/sardinia/index.htm	1

14 26%

	Entrance Points	Visits
1.	/search_book_item.asp	3
2.	-	0
3.	-	0
4.	-	0
5.	-	0

Submit Order Click

Improved layout

Removed

Improved layout

2.23%

Goal page

Installation

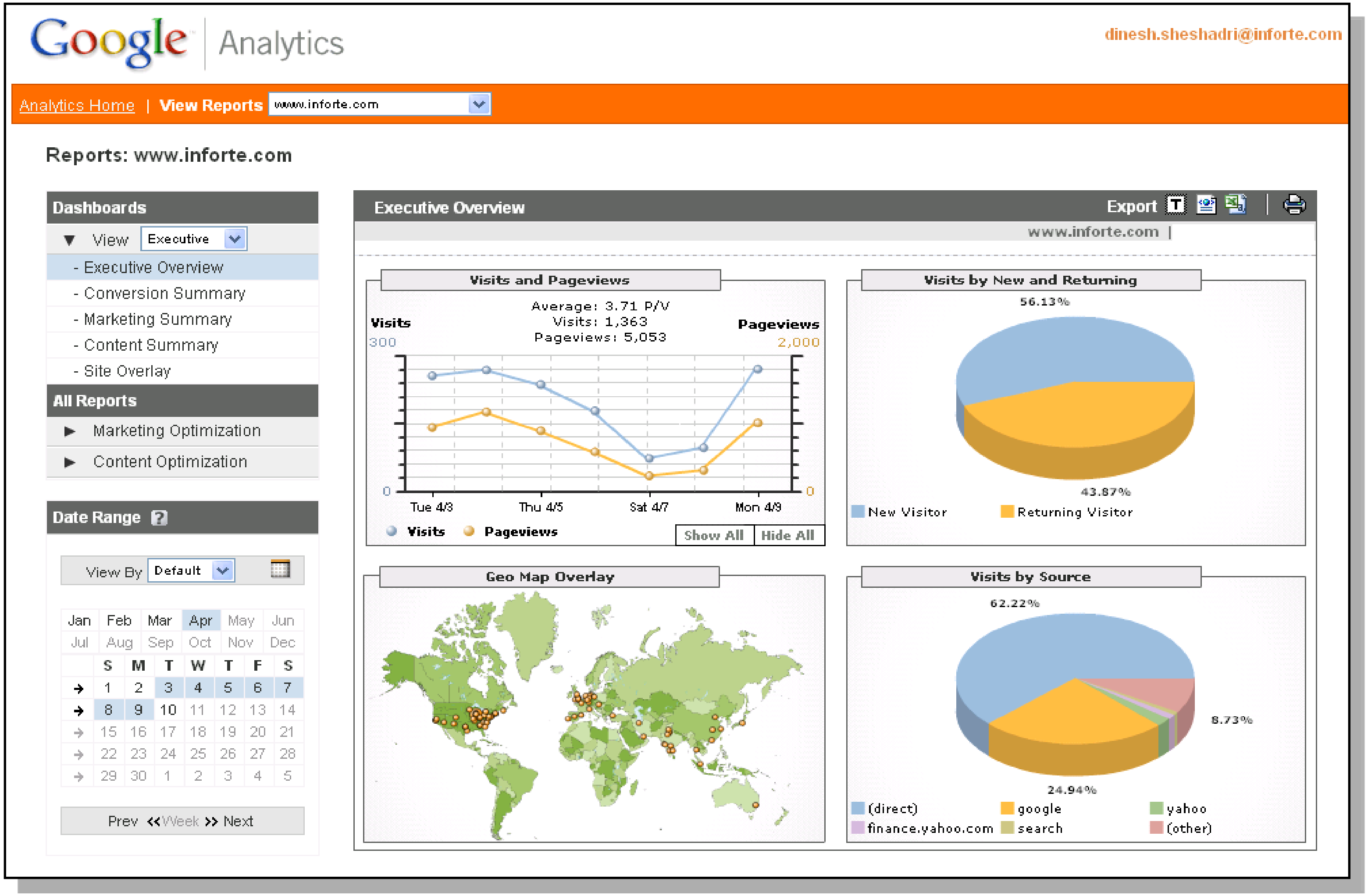
- Get a free account at http://www.google.com/analytics/
- Fill in the account profile and obtain a tracking code
- Add the tracking code to your web pages (which is invisible from your web clients)
-
```
<script type="text/javascript">
var gaJsHost = (("https:" == document.location.protocol) ? "https://ssl." : "http://www.");
document.write(unescape("%3Cscript src='" + gaJsHost + "google-analytics.com/ga.js'
type='text/javascript'%3E%3C/script%3E"));
</script>
```
-
```
<script type="text/javascript">
var pageTracker = _gat._getTracker("UA-xxxxxx-x");
pageTracker._initData();
pageTracker._trackPageview();
</script>
```

Tracking code

Alternatively, these JavaScript code (called "tags" in Google's terms) can be added through Google Tag Manager (GTM), an online tool for non-developers to add "tags" for GA, AdWords, etc. Refer to the post-class references for detail.

- Once this is set, Google will start to count the number of access when your clients retrieve your pages
  - Your clients would also retrieve those hidden objects from Google at the same time, so Google would know and log their accesses
  - Like storing another web server log
- Goals, Funnels, and other filters can be defined and customized
  - Concise documentation is provided at GA website. Easy to follow.
- We will have a short demo on Google Analytics after this lecture

Executive overview report



Analytics

dinesh.sheshadri@inforte.com

Analytics Home | View Reports www.inforte.com

Reports: www.inforte.com

Dashboards

▼ View Executive

- Executive Overview
- Conversion Summary
- Marketing Summary
- Content Summary
- Site Overlay

All Reports

▶ Marketing Optimization

▶ Content Optimization

Date Range

View By Default

Jan	Feb	Mar	Apr	May	Jun
Jul	Aug	Sep	Oct	Nov	Dec

	S	M	T	W	T	F	S
→	1	2	3	4	5	6	7
→	8	9	10	11	12	13	14
→	15	16	17	18	19	20	21
→	22	23	24	25	26	27	28
→	29	30	1	2	3	4	5

Prev << Week >> Next

Executive Overview

Export

www.inforte.com |

Visits and Pageviews

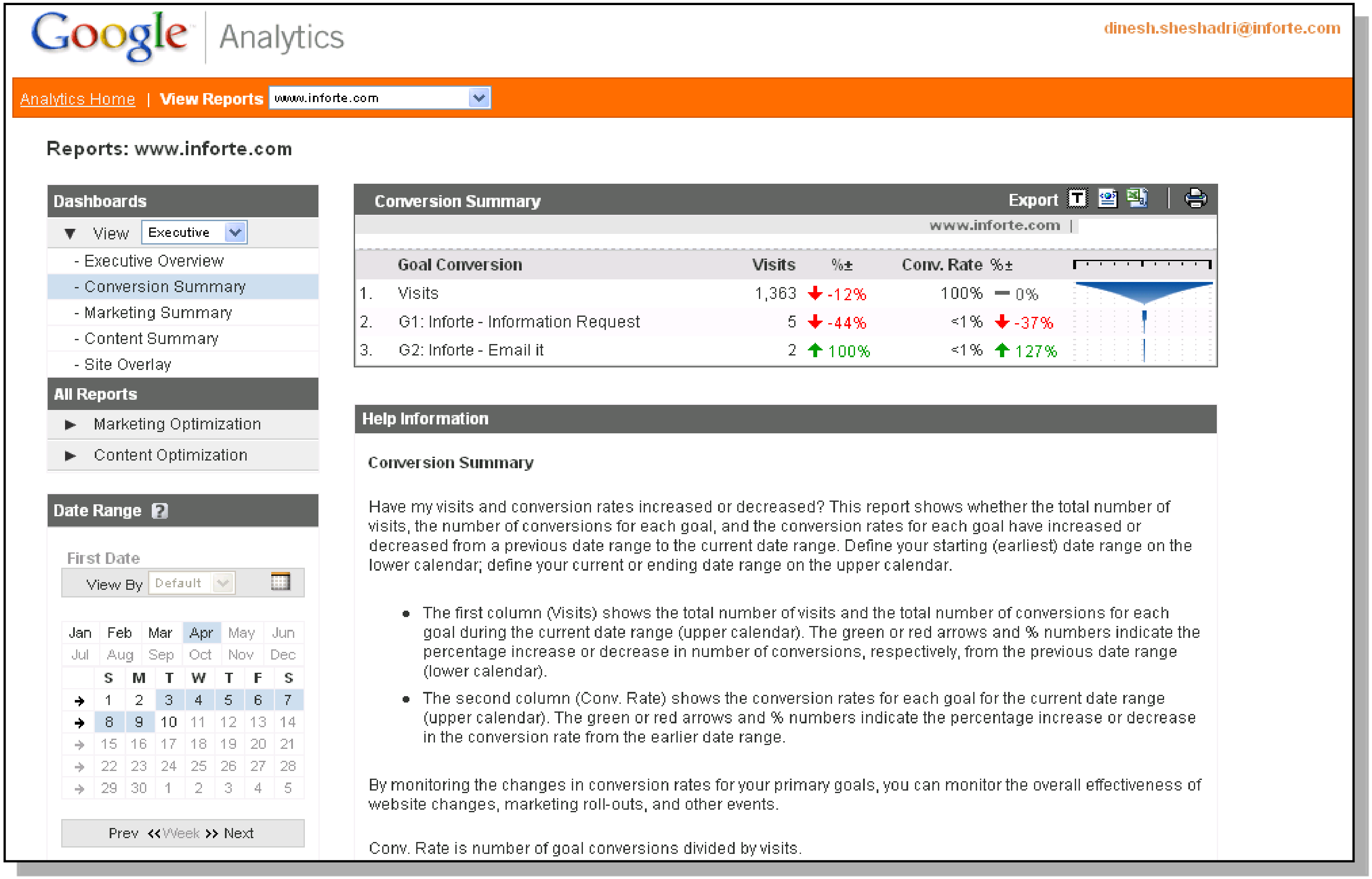
Average: 3.71 P/V
Visits: 1,363
Pageviews: 5,053

Visits by New and Returning

Geo Map Overlay

Visits by Source

Conversion summary report



Jan	Feb	Mar	Apr	May	Jun
Jul	Aug	Sep	Oct	Nov	Dec

	S	M	T	W	T	F	S
→	1	2	3	4	5	6	7
→	8	9	10	11	12	13	14
→	15	16	17	18	19	20	21
→	22	23	24	25	26	27	28
→	29	30	1	2	3	4	5

Conversion Summary

www.inforte.com |

	Goal Conversion	Visits	%±	Conv. Rate	%±
1.	Visits	1,363	↓ -12%	100%	— 0%
2.	G1: Inforte - Information Request	5	↓ -44%	<1%	↓ -37%
3.	G2: Inforte - Email it	2	↑ 100%	<1%	↑ 127%

Help Information

Conversion Summary

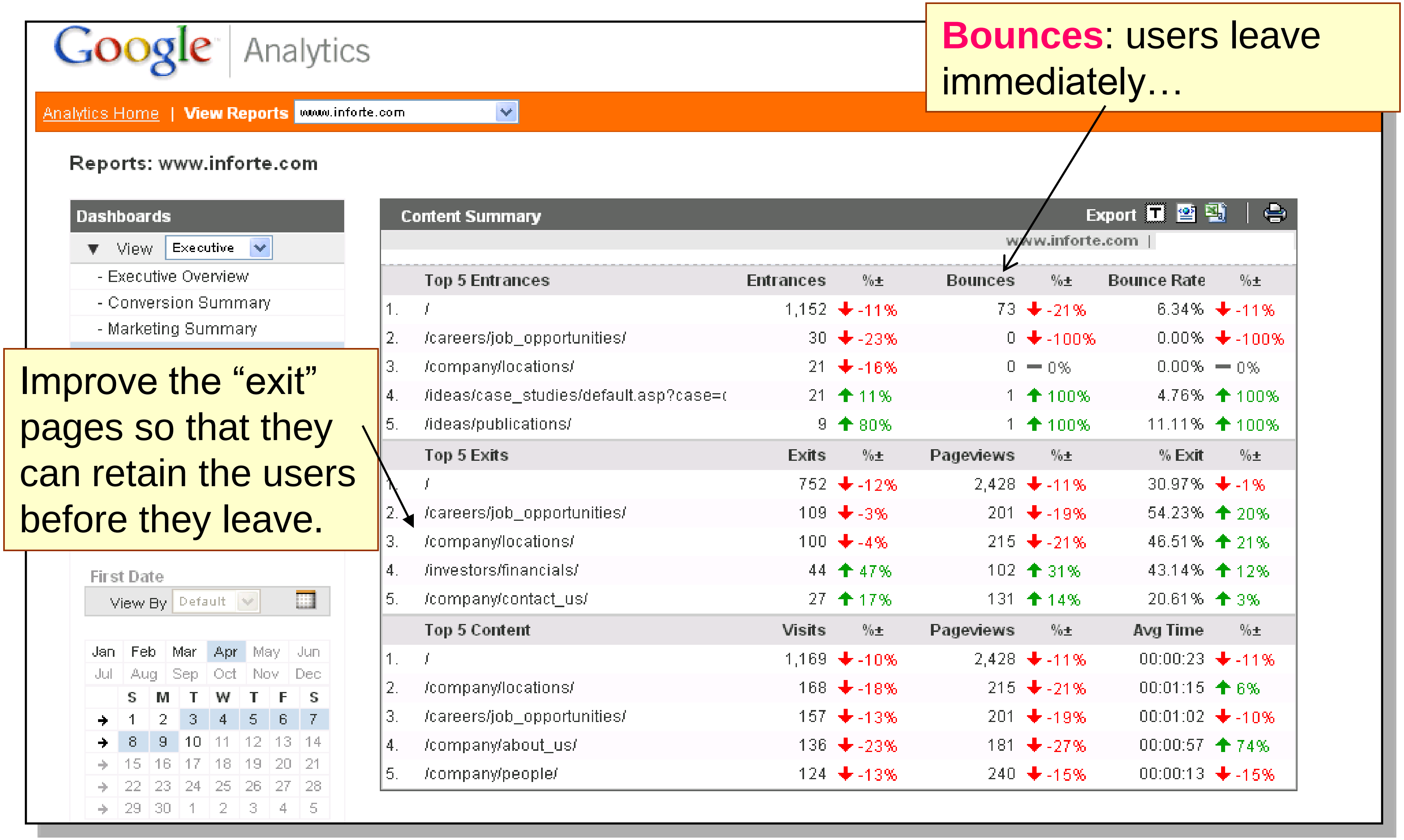
Have my visits and conversion rates increased or decreased? This report shows whether the total number of visits, the number of conversions for each goal, and the conversion rates for each goal have increased or decreased from a previous date range to the current date range. Define your starting (earliest) date range on the lower calendar; define your current or ending date range on the upper calendar.

- The first column (Visits) shows the total number of visits and the total number of conversions for each goal during the current date range (upper calendar). The green or red arrows and % numbers indicate the percentage increase or decrease in number of conversions, respectively, from the previous date range (lower calendar).
- The second column (Conv. Rate) shows the conversion rates for each goal for the current date range (upper calendar). The green or red arrows and % numbers indicate the percentage increase or decrease in the conversion rate from the earlier date range.

By monitoring the changes in conversion rates for your primary goals, you can monitor the overall effectiveness of website changes, marketing roll-outs, and other events.

Conv. Rate is number of goal conversions divided by visits.

Content summary report



Reports: www.inforte.com

Dashboards

View: Executive

- Executive Overview
- Conversion Summary
- Marketing Summary

First Date

View By: Default

Jan	Feb	Mar	Apr	May	Jun
Jul	Aug	Sep	Oct	Nov	Dec

	S	M	T	W	T	F	S
→	1	2	3	4	5	6	7
→	8	9	10	11	12	13	14
→	15	16	17	18	19	20	21
→	22	23	24	25	26	27	28
→	29	30	1	2	3	4	5

Content Summary

www.inforte.com |

	Top 5 Entrances	Entrances	%±	Bounces	%±	Bounce Rate	%±
1.	/	1,152	↓ -11%	73	↓ -21%	6.34%	↓ -11%
2.	/careers/job_opportunities/	30	↓ -23%	0	↓ -100%	0.00%	↓ -100%
3.	/company/locations/	21	↓ -16%	0	— 0%	0.00%	— 0%
4.	/ideas/case_studies/default.asp?case=(	21	↑ 11%	1	↑ 100%	4.76%	↑ 100%
5.	/ideas/publications/	9	↑ 80%	1	↑ 100%	11.11%	↑ 100%

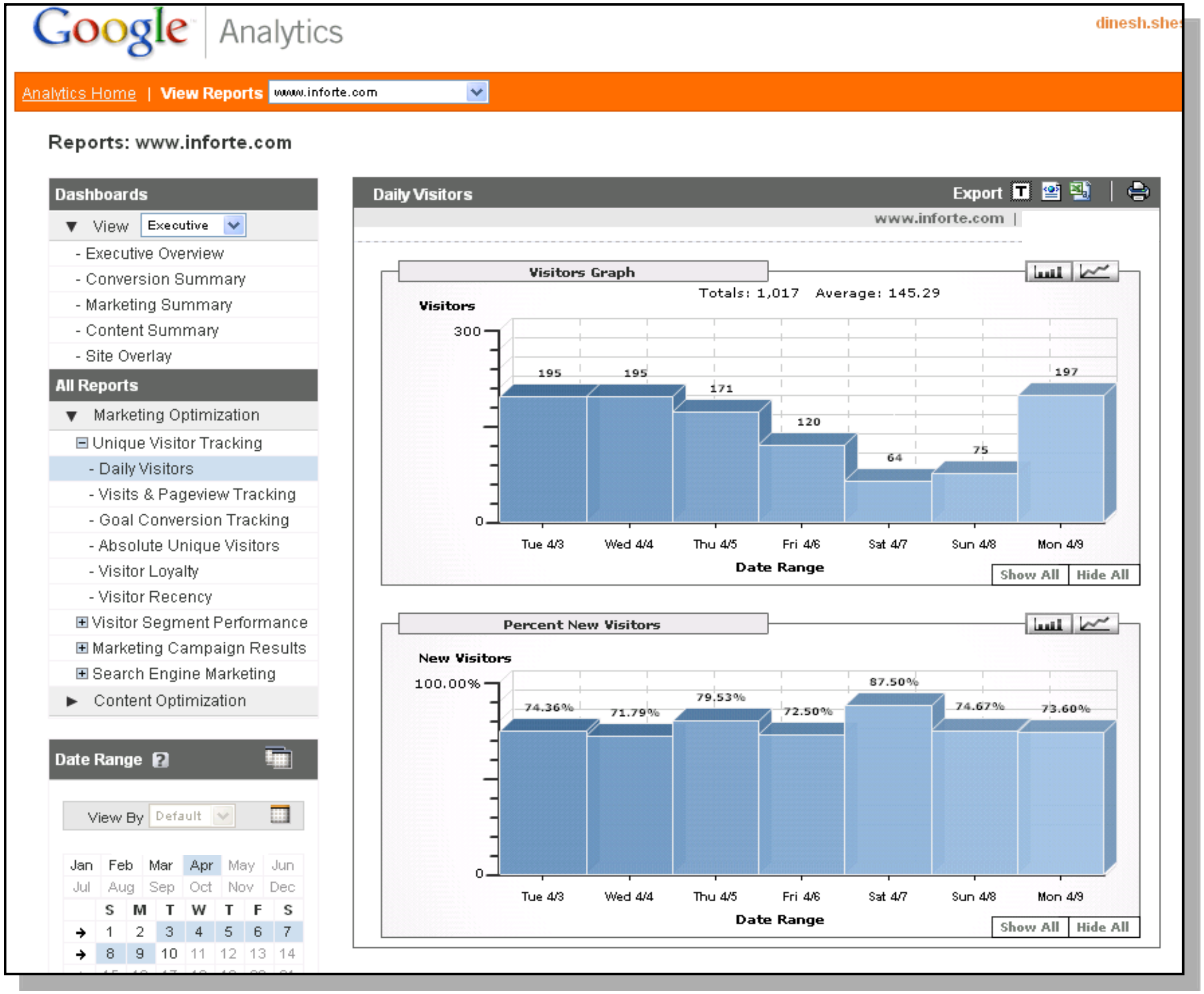
	Top 5 Exits	Exits	%±	Pageviews	%±	% Exit	%±
1.	/	752	↓ -12%	2,428	↓ -11%	30.97%	↓ -1%
2.	/careers/job_opportunities/	109	↓ -3%	201	↓ -19%	54.23%	↑ 20%
3.	/company/locations/	100	↓ -4%	215	↓ -21%	46.51%	↑ 21%
4.	/investors/financials/	44	↑ 47%	102	↑ 31%	43.14%	↑ 12%
5.	/company/contact_us/	27	↑ 17%	131	↑ 14%	20.61%	↑ 3%

	Top 5 Content	Visits	%±	Pageviews	%±	Avg Time	%±
1.	/	1,169	↓ -10%	2,428	↓ -11%	00:00:23	↓ -11%
2.	/company/locations/	168	↓ -18%	215	↓ -21%	00:01:15	↑ 6%
3.	/careers/job_opportunities/	157	↓ -13%	201	↓ -19%	00:01:02	↓ -10%
4.	/company/about_us/	136	↓ -23%	181	↓ -27%	00:00:57	↑ 74%
5.	/company/people/	124	↓ -13%	240	↓ -15%	00:00:13	↓ -15%

Bounces: users leave immediately…

Improve the “exit” pages so that they can retain the users before they leave.

Daily traffic



Reports: www.inforte.com

Daily Visitors

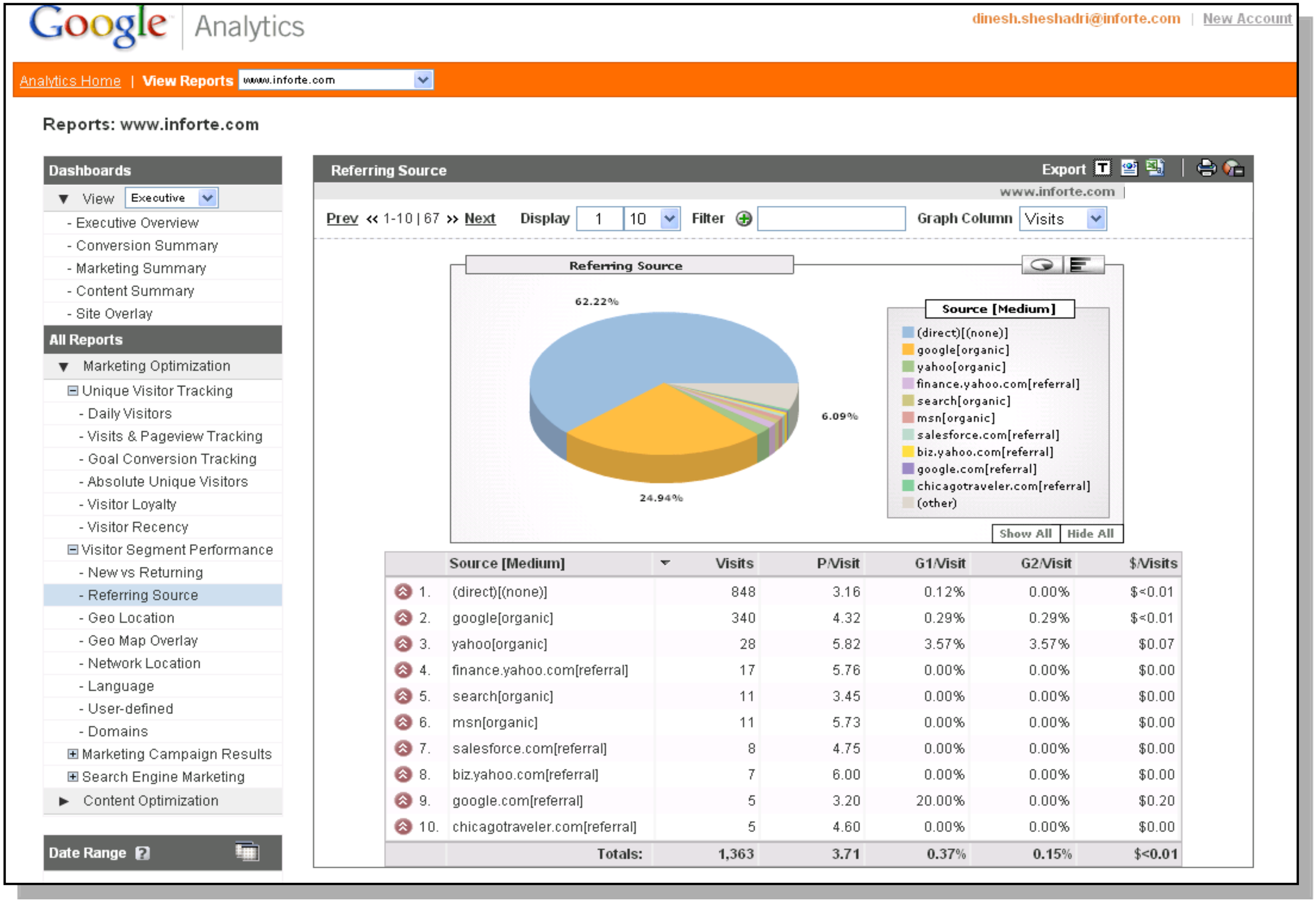
www.inforte.com |

Visitors Graph

Totals: 1,017 Average: 145.29

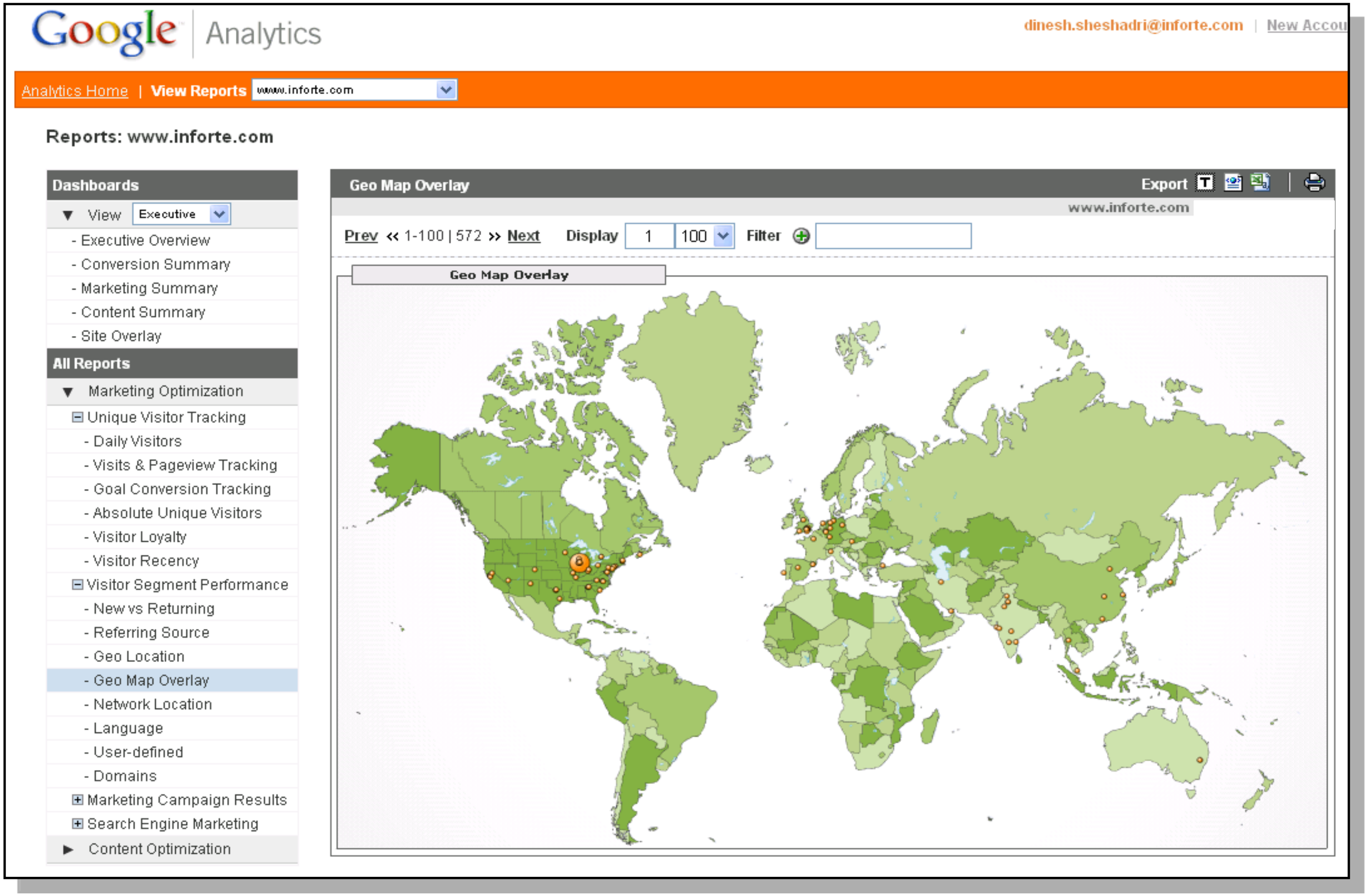
Percent New Visitors

Referring sources



	Source [Medium]	Visits	P/Visit	G1/Visit	G2/Visit	$/Visits
1.	(direct)[(none)]	848	3.16	0.12%	0.00%	$<0.01
2.	google[organic]	340	4.32	0.29%	0.29%	$<0.01
3.	yahoo[organic]	28	5.82	3.57%	3.57%	$0.07
4.	finance.yahoo.com[referral]	17	5.76	0.00%	0.00%	$0.00
5.	search[organic]	11	3.45	0.00%	0.00%	$0.00
6.	msn[organic]	11	5.73	0.00%	0.00%	$0.00
7.	salesforce.com[referral]	8	4.75	0.00%	0.00%	$0.00
8.	biz.yahoo.com[referral]	7	6.00	0.00%	0.00%	$0.00
9.	google.com[referral]	5	3.20	20.00%	0.00%	$0.20
10.	chicagotraveler.com[referral]	5	4.60	0.00%	0.00%	$0.00
	Totals:	1,363	3.71	0.37%	0.15%	$<0.01

Geo map overlay



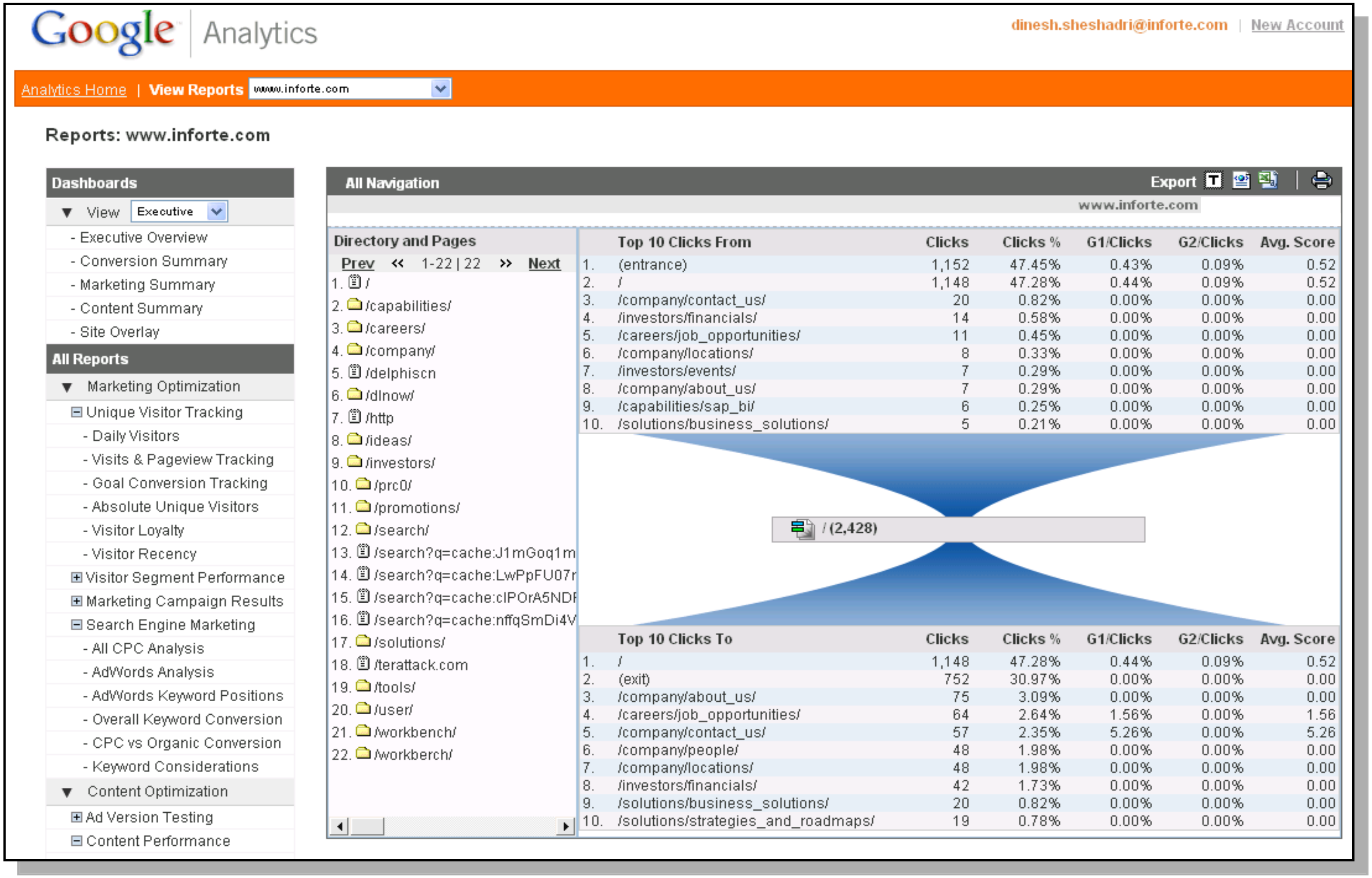
Reports: www.inforte.com

View: Executive

Prev « 1-100 | 572 » Next

Display 1 100

All navigation, and other reports



Google Analytics
dinesh.sheshadri@inforte.com | New Account
Analytics Home | View Reports www.inforte.com

Reports: www.inforte.com

Directory and Pages

Prev « 1-22 | 22 » Next

1. /
2. /capabilities/
3. /careers/
4. /company/
5. /delphiscn
6. /dlnow/
7. /http
8. /ideas/
9. /investors/
10. /prc0/
11. /promotions/
12. /search/
13. /search?q=cache:J1mGoq1m
14. /search?q=cache:LwPpFU07r
15. /search?q=cache:clPOrA5NDI
16. /search?q=cache:nffqSmDi4V
17. /solutions/
18. /terattack.com
19. /tools/
20. /user/
21. /workbench/
22. /workberch/

	Top 10 Clicks From	Clicks	Clicks %	G1/Clicks	G2/Clicks	Avg. Score
1.	(entrance)	1,152	47.45%	0.43%	0.09%	0.52
2.	/	1,148	47.28%	0.44%	0.09%	0.52
3.	/company/contact_us/	20	0.82%	0.00%	0.00%	0.00
4.	/investors/financials/	14	0.58%	0.00%	0.00%	0.00
5.	/careers/job_opportunities/	11	0.45%	0.00%	0.00%	0.00
6.	/company/locations/	8	0.33%	0.00%	0.00%	0.00
7.	/investors/events/	7	0.29%	0.00%	0.00%	0.00
8.	/company/about_us/	7	0.29%	0.00%	0.00%	0.00
9.	/capabilities/sap_bi/	6	0.25%	0.00%	0.00%	0.00
10.	/solutions/business_solutions/	5	0.21%	0.00%	0.00%	0.00

/ (2,428)

	Top 10 Clicks To	Clicks	Clicks %	G1/Clicks	G2/Clicks	Avg. Score
1.	/	1,148	47.28%	0.44%	0.09%	0.52
2.	(exit)	752	30.97%	0.00%	0.00%	0.00
3.	/company/about_us/	75	3.09%	0.00%	0.00%	0.00
4.	/careers/job_opportunities/	64	2.64%	1.56%	0.00%	1.56
5.	/company/contact_us/	57	2.35%	5.26%	0.00%	5.26
6.	/company/people/	48	1.98%	0.00%	0.00%	0.00
7.	/company/locations/	48	1.98%	0.00%	0.00%	0.00
8.	/investors/financials/	42	1.73%	0.00%	0.00%	0.00
9.	/solutions/business_solutions/	20	0.82%	0.00%	0.00%	0.00
10.	/solutions/strategies_and_roadmaps/	19	0.78%	0.00%	0.00%	0.00

Search engine optimizations

How SEO works

- The goal of the search engine is to bring up the most relevant websites for human

- Search engine optimization (SEO) is the process of applying techniques for improving the ranking of websites in search engine results

- In order to identify the most relevant websites, search engines use a set of algorithms to calculate the relevance and ranking of websites w.r.t. search keywords

- Most search engines calculate the ranking of a webpage based on the following factors:
  - The number of external links to the page (“recommendations”)
  - Page title
  - Content (i.e., the keywords)
  - The internal link structure

- => A very high-level summary of SEO – a website should rank high if:
  - There are many external links pointing to it
  - There are relevant keywords/content in the titles and the pages
  - The entire website is well organized (linked) internally

Techniques that improve ranking

Four categories:

1. URL-related
2. Text/content-related
3. Links-related
4. HTML code-related

Some notes:

- Search engines change their ranking algorithms from time to time, a technique worked in the past may not work in the future
  - Prevent "manipulations" where some websites are "engineered" to dominate search results (e.g., content farms)
  - New websites can have chance to appear instead of always showing the same, old but popular websites
- **However**: many of the following techniques are not only for SEO, but also good practices in presenting content to human.
- So, naturally, it is good to follow them. If a technique is good to human readers, it should be welcome by search engines no matter how they change their calculations – since the intended audience of search engines is always human

1.1. Domain name

- Use relevant keywords as part of the domain name whenever possible
  - E.g. HKholidays.com (probably for a travel website related to HK)

- Register other domain extensions to protect your names (and your unique keywords)
  - Register the .net, .org, and other extensions

- If you have multiple domains (e.g., HKHolidays.com, HKHolidays.net, etc.), don't point these domains to the same server IP. Instead, use HTTP redirections (i.e., the 301 response code) instead
  - E.g., all requests sent to the .net domain would be redirected to the .com domain
  - A redirection tells a search engine that the website has been permanently moved
    - A technique approved by the search engines
    - Can pass (most of) the "ranking power" of the original page/domain to the target page/domain
    - http://www.webconfs.com/how-to-redirect-a-webpage.php

1.2. Folder and file names

- Use keywords in folder/file names relevant to their content
  - A page named SEO-Dos-and-Donts.html is better than page17.html
- Use keywords in subfolder names that relate to the subject matter of the division
  - http://www.domain.com/SEOWhitePapers/ is better than http://www.domain.com/folder7
- Minimize use of subfolders
  - The less complicated the structure of the website, the more likely it is to be completely "crawled" (by search engines' robots) and indexed

1.3. Non-Indexable characters

- Avoid non-indexable characters in URLs

  = , ?, etc.

- For query strings, keep them simple
  - ?category=seo is better than ?catid=1&subcatid=15&lastpage=whitepaper

2.1. Text: page content

- Write content **with relevant and important keywords** in mind
  - Use page headings and subheadings (<h1>, <h2>, etc.) where appropriate – size does matter
- Add new content regularly to the web site
  - The more frequently content is added to a web site, the more frequently search engine robots come to that website again
  - A frequently-updated website is more likely to draw users' continual attention – so search engines may rank them higher
- Don't put important keyboards *only* in images because the search engines can't recognize them
- Don't create machine-generated pages with fake content and artificially-inserted keywords
  - This is easily detected (and penalized) by search engines
- Don't repeat/duplicate content across pages
  - Search engines generally avoid/penalize duplicated contents (as a penalty to those content farms/link farms)

2.2. Keywords and text frequency

- Research and find keywords and phrases that are relevant to the website
  - E.g., use tools such as Google AdWords Keyword Planner

- Highly general/popular keywords usually means more readers' interest **and** also competition (in ranking), and vice versa
  - "Mobile phones" vs. "iPhone users, Hong Kong"
  - A tradeoff that every website has to consider

- Find ways to **naturally** incorporate keywords in the website
  - Make the content keyword-rich naturally
  - Review the existing pages and look for opportunities to enhance (or expand) it with keywords
  - E.g., include a personal/corporate blog, newsletters, product reviews, white papers, etc.

Keywords and text frequency

- Don’t use “cloaking”
  - Serve one page to search engines and another to human
  - Or, hide keywords that are targeted to search engines (not to human) with very small font size or other CSS tricks
  - Or, hide keywords behinds objects
  - => Can be detected easily and penalized

- Don’t add keywords that are not related to the website for the purpose of driving traffic
  - Keywords should be as focused and relevant as possible

- In general, meaningless contents prohibit external linking, which in turn would greatly affect ranking
  - The “net” effect (of using meaningless content) would always be a lower ranking instead of higher

2.3. Localization / geo-targeting

- Add geocentric terms and keywords to target local areas
- Add local content per area
  - Especially when serving multiple regions - each region needs local, relevant information
  - Search engines do the same as well, i.e., they search for local info (e.g., HK-specific keywords) for a localized search page (e.g., google.com.hk)
- Don’t generate pages that are identical except for the name of the city, etc.
  - Again, duplicated content can be penalized

3.1. Links

- Use links with meaningful textual descriptions
  - E.g.,
    <a href=“…”>a white paper on SEO written by N. Yim</a>
    is better than
    <a href=“…”>a white paper</a>
  - Search engines are less interested in graphical links

- Link keywords to other relevant pages within the website
  - “**Internal linking**”
  - Provides strategic crawling information for search engines

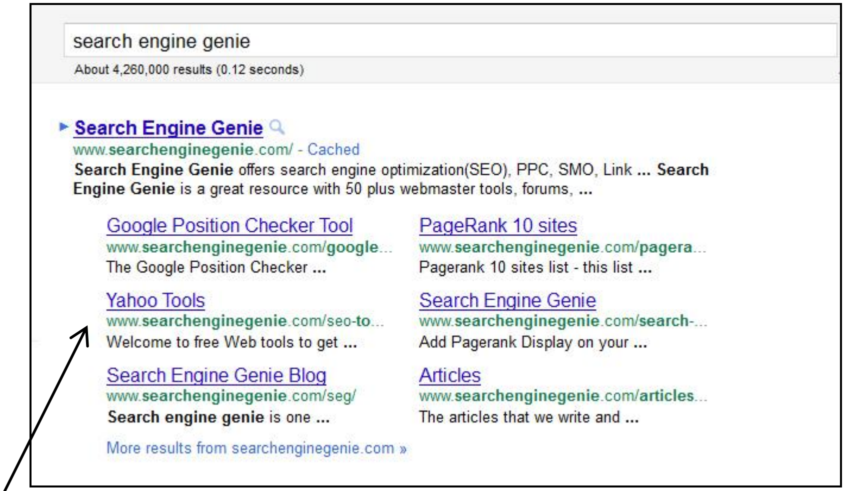
- Have every page of the website accessible through a link somewhere else, either internal or external to the domain
  - So that they can be found by search engines and indexed

3.2. Search engine submission

- Submit by hand *once* to each major search engine, e.g.,:
  - □ https://www.google.com/webmasters/
  - □ https://www.bing.com/webmasters/

- Don’t submit every web page in your website individually to a search engine
  - □ **URLs are better when discovered by a search engine**

- Don’t use automated submission programs – they are more likely to be detected, ignored or penalized



search engine genie

About 4,260,000 results (0.12 seconds)

Search Engine Genie
www.searchenginegenie.com/ - Cached
Search Engine Genie offers search engine optimization(SEO), PPC, SMO, Link ... **Search Engine Genie** is a great resource with 50 plus webmaster tools, forums, ...

Google Position Checker Tool
www.searchenginegenie.com/google...
The Google Position Checker ...

PageRank 10 sites
www.searchenginegenie.com/pagera...
Pagerank 10 sites list - this list ...

Yahoo Tools
www.searchenginegenie.com/seo-to...
Welcome to free Web tools to get ...

Search Engine Genie
www.searchenginegenie.com/search-...
Add Pagerank Display on your ...

Search Engine Genie Blog
www.searchenginegenie.com/seg/
Search engine genie is one ...

Articles
www.searchenginegenie.com/articles...
The articles that we write and ...

More results from searchenginegenie.com »

- Use XML Sitemaps (sitemaps.org) to inform search engines the list of pages that are open for crawling
  - □ Can also help search engines to return Sitelinks to your website (but no guarantees)
  - □ Supported by Google, Yahoo and Microsoft
  - □ Prepare a sitemap.xml (by using an online tools, or make one manually), then include it in robots.txt to be placed in the root folder of your website, e.g.,

```
Sitemap: http://domain.com/sitemap.xml
```

3.3. In-bound link (IBL) development

- Look for industry-related, authority websites to acquire links from
  - E.g., if your website is technology-related, it would be great if a related article at CNET points to one of your pages

- Document link development progress and regularly invest time for further growth

- Use appropriate keywords (which describe what your website is about) for the anchor text on external links

- Organic in-bound links (i.e., links added by human, e.g., in an article) are much better than those generated, so
  - Don't buy links for the purpose of IBL development
  - Don't participate in link exchanges or link farms
    - They have no purpose other than providing advertisements and links (i.e., they are quite likely to be penalized)
  - Don't participate in "forum spam" or "guestbook spam"
    - A technique that automatically posts links to your website on other web pages in the "comment" section, e.g., blogs, YouTube.com, etc.
  - If you subscribe to any services / tools that claim to help IBL development, do check carefully what they really do

4.1. Optimize the first 100 lines of code

- Search engines (those “robots”) are designed to behave like human:
  - They tend to focus on the “top” part of web pages
  - => Optimizing the relevance of content at the **top** of pages is crucial
- Instead of including your javascript and CSS style definitions in the <head> section, we can make them external
  - <script language="JavaScript" type="text/javascript" src=“insert-JavaScript-file-name.js"></script>
  - <link rel="stylesheet" href=“insert-style-file-name.css">
- Leave it for <meta> tags and meaningful content

4.2. Meta tags

- The keyword meta tag helps the search engines understand which keywords are most related to a page
  - But: only use it for important, relevant keywords
  - Limit the keywords in the meta tag to 10 or below

- Google doesn't seem to consider the “keyword” meta tags anymore. But some other search engines consider.

- Don’t stuff meta tags with irrelevant/repeated keywords
  - This abuse has caused the devaluation of meta tags since long time ago

- For example, the following website tries to be ranked high for "tents" – a typical misuse of the keyword meta tag:

```
<meta name=“keywords" content="tents, TENTS, Tents, tents tents tent supplies,
tents, tents tent, tent, Tent, TENTS, tents, Tents,tents, TENTS, tent, tent,
Tent, TENTS, tents, Tents,tents, TENTS, Tents, tents tents tent supplies,
tents, tents tent, tent, Tent, TENTS, tents, Tents
tents, TENTS, Tents, tents tents tent supplies, tents, tents tent, tent, Tent,
TENTS, tents, Tents">
```

- Bing penalizes such usage

4.3. Title tag

- Write short titles with the most important keywords or phrases in them
  - □ One good format is “Important Keyword Phrases – Company Name”
  - □ E.g., “Product catalogue – ABC Mobile Phones”
  - □ E.g., “Course outline – Website Engineering, ECom-IComp, HKU”

For example, this one is 52 characters long

- Make the Title concise
  - □ Ideally 20-50 characters
  - □ Should not be too long (e.g., 100+ characters)
  - □ Don’t include misleading/repeated keywords
  - □ Don’t use HTML markup, e.g., <b>, <i>, etc.

- Write unique titles for each web page based on the content
  - □ => each web page would have the opportunity to be ranked for valuable keywords
  - □ => a larger “keyword space”

- Don’t use multiple title tags
  - □ This does not work and may prevent the correct title from being indexed

4.4. Image Alt Tag

- Image Alt Tag:
  - **`<img src= "image.jpg" alt="meaningful and descriptive text here">`**

- Can be used to include meaningful keywords that describe the picture and the subject matter of the page

- E.g., set the alt tag of the company logo to be the company name and slogan

- E.g., set the alt tag of product photos as product names or short descriptions
  - **For image search**

- Or use the figcaption tag for figures in HTML5

4.5. Other SEO considerations

- Don’t use frames
    - The website structure with frames is difficult to be understood

- Don’t use graphics/Adobe Flash for an entire website – they are difficult to be understood by search engines

- It usually takes time (a month to 1+ years) to do good SEO (e.g., to build up inbound-links)
    - Not a “one-off” investment or process
    - Regular efforts (i.e., investment) needed

Some SEO resources

- Useful and updated articles and experience sharing:
  - http://www.searchenginewatch.com
  - http://www.searchengineguide.com

Post class self-learning resources

- Post-class readings:
  - □ Take a brief look at Chrome DevTools or Firefox Developer Tools
  - □ Take a brief look at AWStats, Piwik/Matomo, and Google Analytics
  - □ Google Analytics for Beginners
  - □ List of web analytics software
  - □ Search Engine Watch and Search Engine Guide

- Reference materials (e-books on Google Analytics and SEO):
  - □ F. Alhlou et al. Google Analytics breakthrough: from zero to business impact. John Wiley & Sons. 2016.
  - □ A. Shenoy et al. Introducing SEO: your quick-start guide to effective SEO practices. Apress. 2016.

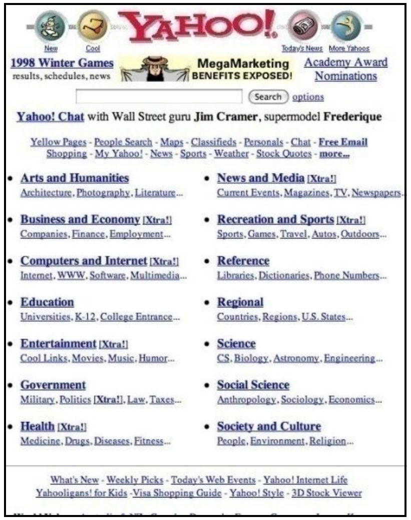
- Two very good books on performance optimizations of websites (not e-books; hardcopy available at HKU Libraries):
  - □ S. Souders. High performance web sites. O'Reilly 2007.
  - □ S. Souders. Even faster web sites. O'Reilly 2009.

- Please see Moodle for the links

Course summary

Summary of scope

Part 1 (lectures 1-3):



1998 Winter Games
results, schedules, news

MegaMarketing
BENEFITS EXPOSED!

Academy Award Nominations

Yahoo! Chat with Wall Street guru **Jim Cramer**, supermodel **Frederique**

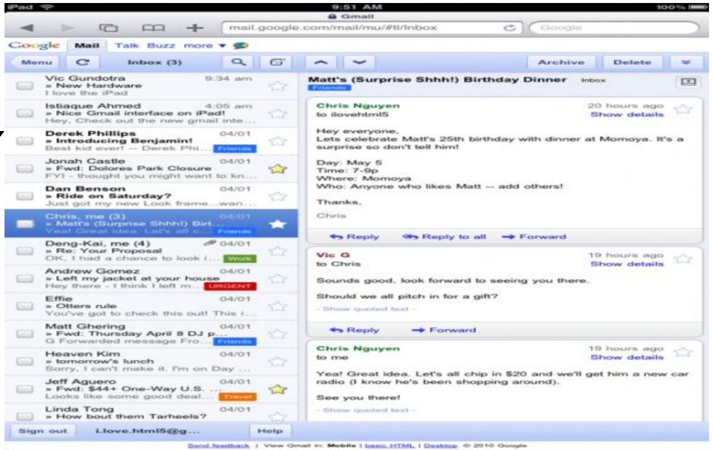
Yellow Pages - People Search - Maps - Classifieds - Personals - Chat - **Free Email**
Shopping - My Yahoo! - News - Sports - Weather - Stock Quotes - **more...**

- **Arts and Humanities**
  Architecture, Photography, Literature...
- **Business and Economy** [Xtra!]
  Companies, Finance, Employment...
- **Computers and Internet** [Xtra!]
  Internet, WWW, Software, Multimedia...
- **Education**
  Universities, K-12, College Entrance...
- **Entertainment** [Xtra!]
  Cool Links, Movies, Music, Humor...
- **Government**
  Military, Politics [Xtra!], Law, Taxes...
- **Health** [Xtra!]
  Medicine, Drugs, Diseases, Fitness...
- **News and Media** [Xtra!]
  Current Events, Magazines, TV, Newspapers...
- **Recreation and Sports** [Xtra!]
  Sports, Games, Travel, Autos, Outdoors...
- **Reference**
  Libraries, Dictionaries, Phone Numbers...
- **Regional**
  Countries, Regions, U.S. States...
- **Science**
  CS, Biology, Astronomy, Engineering...
- **Social Science**
  Anthropology, Sociology, Economics...
- **Society and Culture**
  People, Environment, Religion...

What's New - Weekly Picks - Today's Web Events - Yahoo! Internet Life
Yahooligans! for Kids -Visa Shopping Guide - Yahoo! Style - 3D Stock Viewer

Websites becoming "web apps" -> more sophisticated, lots of interactions with users, "Web 2.0", etc.

Part 2 (lectures 4-6):



Vic Gundotra 9:34 am
» New Hardware
I love the iPad

Istiaque Ahmed 4:05 am
» Nice Gmail interface on iPad!
Hey, Check out the new gmail inte...

Derek Phillips 04/01
» Introducing Benjamin!
Best kid ever! -- Derek Phi... Friends

Jonah Castle 04/01
» Fwd: Dolores Park Closure
FYI - thought you might want to kn...

Dan Benson 04/01
» Ride on Saturday?
Just got my new Look frame...wan...

Chris, me (3) 04/01
» Matt's (Surprise Shhh!) Birt...
Yea! Great idea. Let's all c... Friends

Deng-Kai, me (4) 04/01
» Re: Your Proposal
OK, I had a chance to look i... Work

Andrew Gomez 04/01
» Left my jacket at your house
Hey there - I think I left m... URGENT

Effie 04/01
» Otters rule
You've got to check this out! This i...

Matt Ghering 04/01
» Fwd: Thursday April 8 DJ p...
G Forwarded message Fro... Friends

Heaven Kim 04/01
» tomorrow's lunch
Sorry, I can't make it. I'm on Day ...

Jeff Aguero 04/01
» Fwd: $44+ One-Way U.S. ...
Looks like some good deal... Travel

Linda Tong 04/01
» How bout them Tarheels?

Matt's (Surprise Shhh!) Birthday Dinner Inbox
Friends

Chris Nguyen 20 hours ago
to ilovehtml5 Show details

Hey everyone,
Lets celebrate Matt's 25th birthday with dinner at Momoya. It's a surprise so don't tell him!

Day: May 5
Time: 7-9p
Where: Momoya
Who: Anyone who likes Matt -- add others!

Thanks,
Chris

Reply Reply to all Forward

Vic G 19 hours ago
to Chris Show details

Sounds good, look forward to seeing you there.

Should we all pitch in for a gift?
- Show quoted text -

Reply Forward

Chris Nguyen 19 hours ago
to me Show details

Yea! Great idea. Let's all chip in $20 and we'll get him a new car radio (I know he's been shopping around).

See you there!
- Show quoted text -

Send feedback | View Gmail in: **Mobile** | basic HTML | Desktop © 2010 Google

Integration and interoperability issues -> how to reuse existing & remote data?

Part 3 (lectures 7-9):

YouTube, Gmail, Amazon, online databases, Maps, updated event lists, YOUR websites, ...

The web/cloud(s)

Part 4 (lecture 10):

Optimizations

You have a great website, how to make it loaded fast at users' computers, and most important… popular?

Part 1: basics

- Enabling standards and the standardization process
  - Benefits of standard adoption include: better interoperability, compatibility, community support, proper upgrade path, etc.
  - Important organizations: W3C, IETF, WHATWG, etc.

- Key selection criteria of web technologies
  - Standard-compliance
  - Have a stable, cross-platform reference implementation
  - Have an active community and future development roadmap
  - Vendor-neutral and/or open-source if possible

- A quick review of basic technologies:
  - X/HTML, CSS, JavaScript, DOM, and server-side scripting, etc.
    - An integrated example: the login page of Facebook.com

  - Importance of separation: {document structure, presentation details, behaviors}

  - HTML5 brings new features such as semantic markup, canvas, native audio/video support, new form elements, new support for web applications (e.g., web storage, app cache, geolocation), etc.

  - CSS3 brings new features (e.g., multi-column layout, more visual effects of text/images, etc.) that are otherwise tedious to implement

Part 1: basics

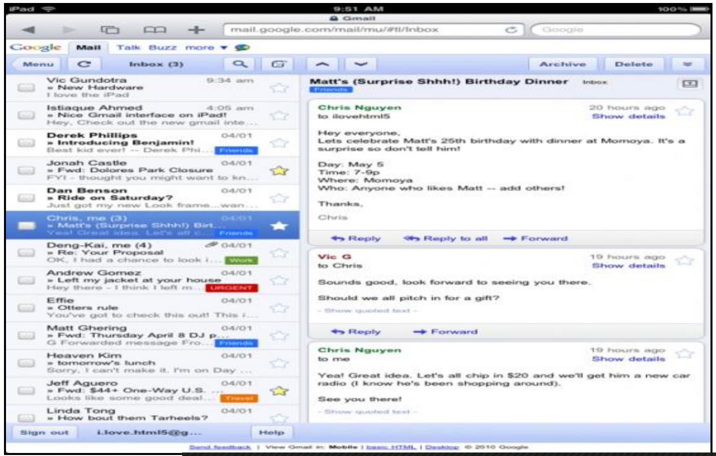
- Web security – common types of attacks and ways to avoid them

- Internationalization/i18n (Unicode; separation of content/data from their validation/presentation logic)

- Responsive web design for supporting mobile/multiple devices
  - Main techniques:
    - Viewport settings
    - CSS media queries
    - Fluid grids
    - Flexible images
  - Progressive enhancement (better) over graceful degradation
    - **“Mobile First”**
    - => Basic content and functionality should be available to mobile browsers, i.e., all browsers
    - 1. Use semantic markup (e.g., <article>, <nav>, etc. in HTML5) to contain all content
    - 2. Place formatting details (for different devices) in external CSS files, which are applied based on device sizes
    - 3. Define behaviors (of HTML elements) in external JavaScript files, applied in the right context

Part 2: web 2.0 applications

- Part 2 is about rapid implementation of web 2.0 sites/applications, and make them maintainable and extensible
- Key concept: leverage third-party code and adopt a well-established design pattern (e.g., MVC)
  - Client-side libraries
  - Server-side frameworks
  - Content management systems (CMS) and Wiki
- Client-side libraries substantially simplify the development of impressive and user-friendly websites
  - E.g., jQuery, Bootstrap, etc.
  - And their plugins and extensions



Vic Gundotra 9:34 am
» New Hardware
I love the iPad

Istiaque Ahmed 4:05 am
» Nice Gmail interface on iPad!
Hey, Check out the new gmail inte...

Derek Phillips 04/01
» Introducing Benjamin!
Best kid ever! -- Derek Phi...

Jonah Castle 04/01
» Fwd: Dolores Park Closure
FYI - thought you might want to kn...

Dan Benson 04/01
» Ride on Saturday?
Just got my new Look frame...wan...

Chris, me (3) 04/01
» Matt's (Surprise Shhh!) Birt...
Yea! Great idea. Let's all c...

Deng-Kai, me (4) 04/01
» Re: Your Proposal
OK, I had a chance to look i...

Andrew Gomez 04/01
» Left my jacket at your house
Hey there - I think I left m...

Effie 04/01
» Otters rule
You've got to check this out! This i...

Matt Ghering 04/01
» Fwd: Thursday April 8 DJ p...
G Forwarded message Fro...

Heaven Kim 04/01
» tomorrow's lunch
Sorry, I can't make it. I'm on Day ...

Jeff Aguero 04/01
» Fwd: $44+ One-Way U.S. ...
Looks like some good deal...

Linda Tong 04/01
» How bout them Tarheels?

Matt's (Surprise Shhh!) Birthday Dinner

Chris Nguyen 20 hours ago
to ilovehtml5
Show details

Hey everyone,
Lets celebrate Matt's 25th birthday with dinner at Momoya. It's a surprise so don't tell him!

Day: May 5
Time: 7-9p
Where: Momoya
Who: Anyone who likes Matt -- add others!

Thanks,
Chris

Vic G 19 hours ago
to Chris
Show details

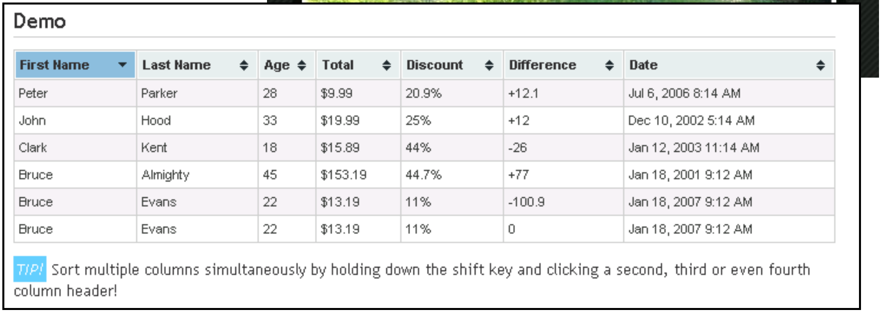
Sounds good, look forward to seeing you there.

Should we all pitch in for a gift?

Chris Nguyen 19 hours ago
to me
Show details

Yea! Great idea. Let's all chip in $20 and we'll get him a new car radio (I know he's been shopping around).

See you there!



Demo

First Name	Last Name	Age	Total	Discount	Difference	Date
Peter	Parker	28	$9.99	20.9%	+12.1	Jul 6, 2006 8:14 AM
John	Hood	33	$19.99	25%	+12	Dec 10, 2002 5:14 AM
Clark	Kent	18	$15.89	44%	-26	Jan 12, 2003 11:14 AM
Bruce	Almighty	45	$153.19	44.7%	+77	Jan 18, 2001 9:12 AM
Bruce	Evans	22	$13.19	11%	-100.9	Jan 18, 2007 9:12 AM
Bruce	Evans	22	$13.19	11%	0	Jan 18, 2007 9:12 AM

TIP! Sort multiple columns simultaneously by holding down the shift key and clicking a second, third or even fourth column header!

Part 2: web 2.0 applications

- Server-side frameworks further extend the "separation" of code and provide convenient supports for implementing complicated application logic
  - Easy form handling, template management, data validation, authentication, object-relational mapping (ORM), etc.
  - E.g., Laravel, Ruby on Rails

- Summary of separations:
  - Client-side: {document structure, presentation details, behaviors}
  - Server-side: {model, view, controller}
  - For i18n: {content, presentation logic, validation logic}

- The concepts of "convention over configuration" and "scaffolding" greatly shorten the code and development time

- CMS/Wiki provides convenient features for supporting "socialization", user communities and collaborations

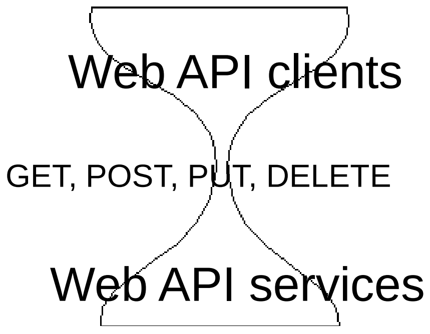
- A "framework-centric approach" to rapid development of web 2.0 websites/apps
  - We have discussed some criteria for choosing the appropriate framework(s)
  - Use the chosen frameworks and tools as much as possible; avoid manual developement
  - Always explore the plugin/extension libraries of the frameworks / CMS before implementing a certain feature

Part 3: interoperability

- Key concept: website/app development can be further simplified by leveraging remote data/service (through web APIs) and cloud services
  - Part 2: achieve the goal (rapid development) by leveraging <u>third-party code/frameworks</u>
  - Part 3: achieve the <u>same goal</u> by leveraging <u>remote data/services</u>

- We can also make our data available to others through web APIs

- Five main architectures of web APIs:
  - Service-orientation (SOAP-based)
  - Resource-orientation (RESTful)
  - RPC-based
  - Feed-based
  - JavaScript



RESTful services

Importance of keeping the web API protocol (the “verbs”) simple for better interoperability (as in REST vs. SOAP)

- Mashups – to combine the output of web APIs and form value-added services
  - An integrated example: Google Maps and Flickr

- Client-side rendering (which use web API extensively) as an emerging technique for website implementation. React as an example.

Part 3: interoperability

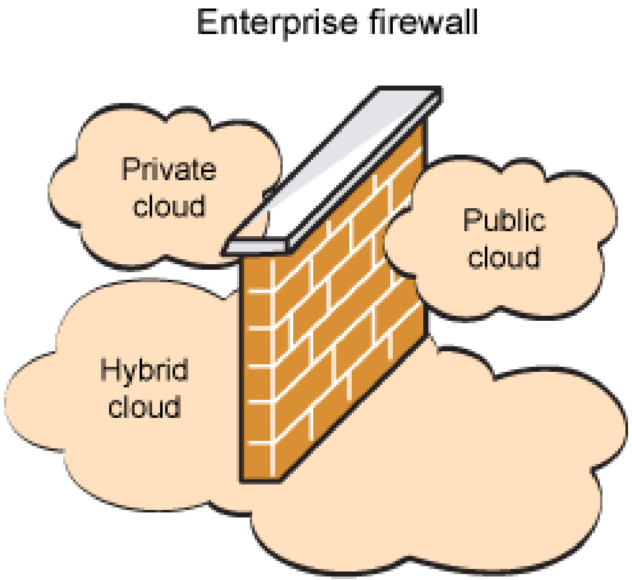
- Cloud computing introduced

- Types of cloud services: SaaS, PaaS, IaaS.

- Advantages: cost effectiveness, platform scalability, data redundancy, advanced security controls, and library supports for rapid development
  - Amazon Web Services
  - Google App Engine

- Cloud computing not only simplifies development, but also streamlines website hosting and management

- Security is a main concern in cloud computing; some issues discussed



Part 4: optimizations (this lecture)

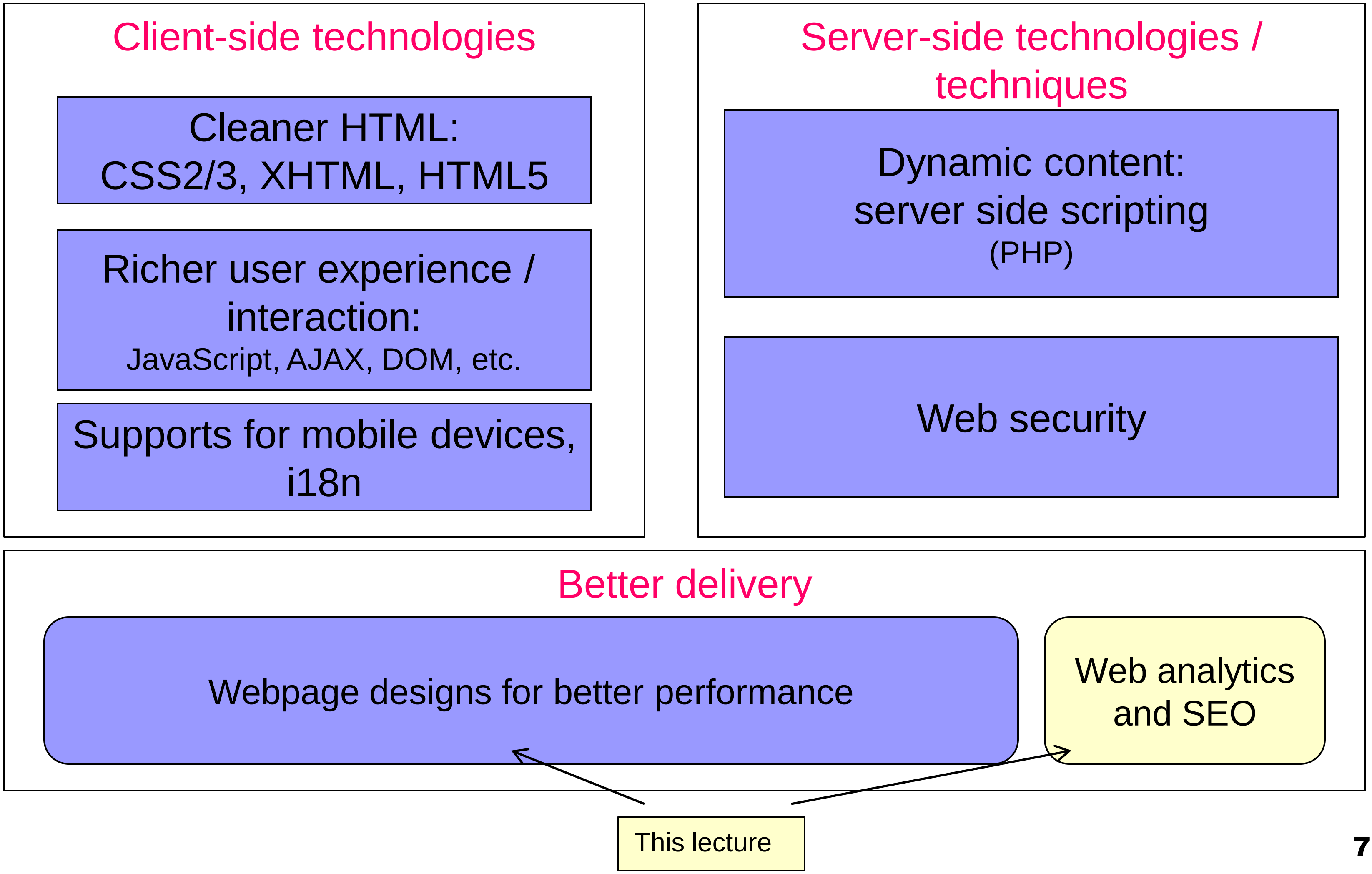
- After we build a website/app, we may want to improve its performance and popularity

- For performance: web page design techniques for faster webpage rendering

- For popularity:
  - □ Techniques for understanding and improving websites' traffic:
    - Log file analysis and hosted analytics services (Google Analytics as an example)
    - Concepts like the clickstreams and conversion funnels are introduced.
  - □ Search engine optimizations:
    - A website's ranking in search engines can be improved by optimizing the URL, text content, (external & internal) hyperlinks and HTML code
    - In general, good approaches to SEO are also good approaches to presenting contents to human readers (a very basic principle in search engines' designs)

Optimizations for performance and popularity are much better to be planned **from Day 1** of the design of a website rather than an afterthought. SEO requires continual efforts.

Technologies covered (web/web 1.0)
Part 1 and Part 4



Client-side technologies

Cleaner HTML:
CSS2/3, XHTML, HTML5

Richer user experience / interaction:
JavaScript, AJAX, DOM, etc.

Supports for mobile devices, i18n

Server-side technologies / techniques

Dynamic content:
server side scripting
(PHP)

Web security

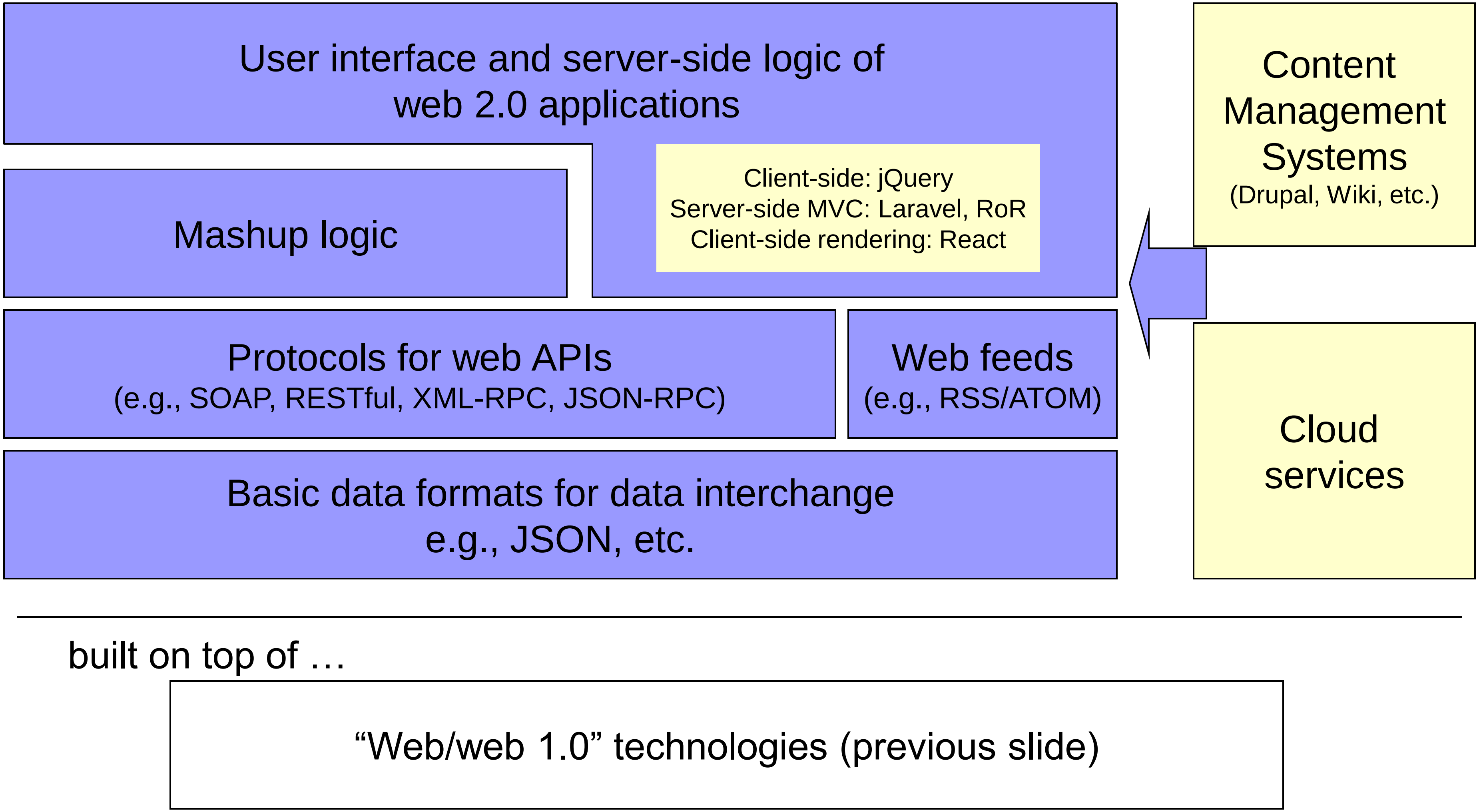
Better delivery

Webpage designs for better performance

Web analytics and SEO

This lecture

Technologies covered (web 2.0): better data reuse, richer functions, and more social interactions
Part 2 and Part 3



User interface and server-side logic of web 2.0 applications

Client-side: jQuery
Server-side MVC: Laravel, RoR
Client-side rendering: React

Mashup logic

Protocols for web APIs
(e.g., SOAP, RESTful, XML-RPC, JSON-RPC)

Web feeds
(e.g., RSS/ATOM)

Basic data formats for data interchange
e.g., JSON, etc.

Content Management Systems
(Drupal, Wiki, etc.)

Cloud services

built on top of …

“Web/web 1.0” technologies (previous slide)

Course objectives revisited

Parts 2 & 3 (frameworks, web APIs, cloud services)

- To introduce the engineering techniques for rapid development of maintainable, extensible, interactive and high-performance websites and web applications.

Part 4: optimizations

Labs, assignment and project

- To provide hands-on experiences with some representative technologies.

- To highlight the importance of open standards/source and standard adoption.

Part 1, and throughout the whole course

Intended learning outcomes

At this point, you should:

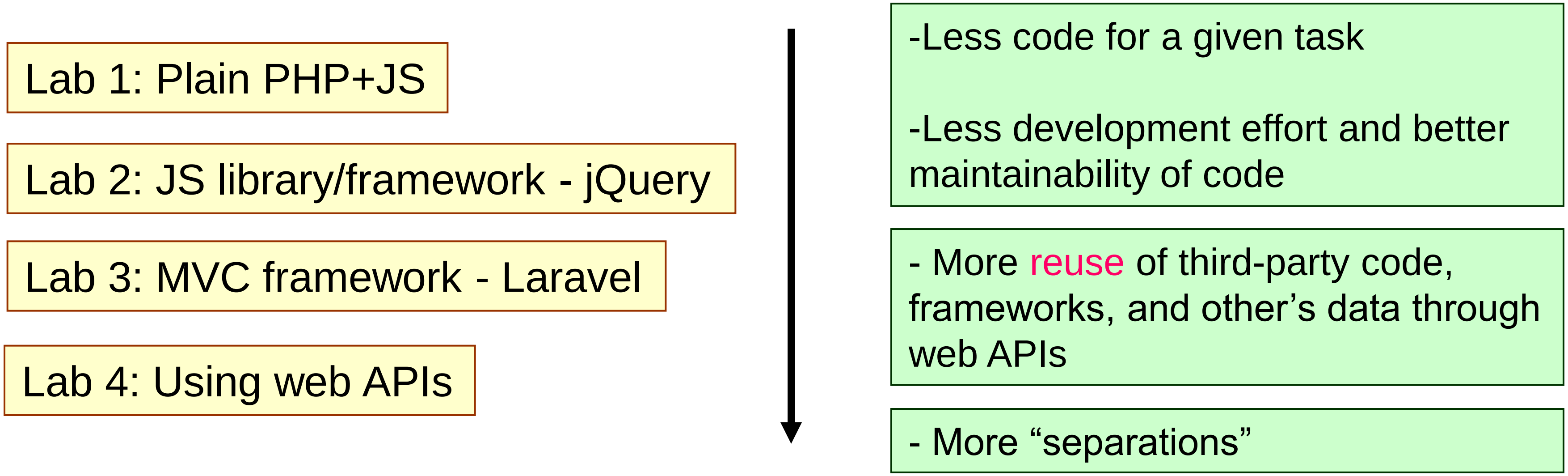
■ Know the standards and the standardization processes of the web	Part 1 and whole course
■ Be able to identify appropriate approaches and tools for building websites based on project needs.	Parts 2 & 3
■ Have hands-on experiences with web development frameworks and libraries.	Labs, assignment/project
■ Know how to simplify website development by using web APIs and cloud computing.	Part 3
■ Be able to speed up the rendering of web pages ■ Know some sustainable approaches to SEO	Part 4

Labs

- The labs are intended to be as simple as possible, but to a depth that one can explore further details easily

- “Model answers” are provided, which serve as a reference for the assignment

- The labs show how website designs can be improved progressively by employing some good engineering practices.



Lab 1: Plain PHP+JS

Lab 2: JS library/framework - jQuery

Lab 3: MVC framework - Laravel

Lab 4: Using web APIs

-Less code for a given task

-Less development effort and better maintainability of code

- More reuse of third-party code, frameworks, and other’s data through web APIs

- More “separations”

Course coverage

- "Breadth" over "depth" – we outline and focus on the big picture
- Website engineering is a broad subject; virtually impossible to cover everything in-depth.
  - And probably not useful to do so because technologies change so fast
- => We should have a fairly complete coverage of the modern web technologies and some good engineering practices
  - => The "Big picture" that we aim to cover

Further studies

- Post-class readings and reference materials are given at the end of each session:
  - Post-class self-learning resources: more examples similar to what have been introduced in the class; or more technical details
  - Reference materials: complete coverage of a topic for production use
  - A compiled list will be shared on Moodle (in “News and Announcements”) soon – you will receive it through email
- Although we focus on “breadth”, there are topics (perhaps remotely) relating to “website engineering” which have not been covered, e.g.,:
  - Proprietary (or vendor-specific) technologies – concepts are similar to what we have introduced.
  - System administration – universal techniques (e.g., data backup practice, system monitoring, etc.) that are applicable to all kinds of servers
  - Software engineering – project lifecycle and management. Not covered due to time constraints. For further readings, get a good software engineering book that covers web application development (web development is only a special case of software development)
  - Web (or search engine) marketing – might better be covered in an ECOM course.
- Feel free to contact me for information/readings on the above topics.

Always stay informed

- The web is evolving everyday
  - We in 2010 (for example) could hardly imagine how much “web 2.0”, the “cloud” and smartphones could have affected us

- Being able to stay informed is essential for professional web developers and project managers

- A “cost-effective” way to do so is to read related news and articles everyday (e.g., through feeds or something like Flipboard, etc.)
  - E.g., CIO.com, O’Reilly Radar, TechRadar Pro, TechCrunch.com, Readwrite.com, etc.

For prospective web developers: “Reading To Write” – Stephen King

- "If you don't have time to read, you don't have the time (or tools) to write."

- Examples of “reading” (the others’ work) as a basic step to learn
  - When we learn a foreign language, we read newspapers, articles, fictions, etc.
  - When architecture students learn about “design patterns” or styles, they study the design of existing buildings.
  - => Why don’t we do the same when we learn a new *programming* language or design a system or website?

- Reading is <u>also</u> an important learning step for computer programming, system design and engineering in general
  - Learn how a system is written (i.e., implemented) and structured – and why.
  - To learn = to understand + **<u>to be able to apply it</u>**.
  - Through reading the others’ work, we recognize patterns and architectural styles. Steal it and make it our own – then develop our own.

- When you learn about a technology, don’t forget reading - start with a **<u>good</u>** book that has a single, integrated, example. Also look at the design and implementation of **<u>good</u>** open-source software and frameworks.
  - Focus on: <u>how the code works</u>, <u>how it is structured</u>, and <u>why it is structured in that way</u>.

- **Practice makes perfect.**

Final words (for everyone): “Problems” and “solutions” in web development

- Web development has always been driven by “problems” and “solutions”
  - E.g., why do we have libraries/frameworks, web APIs, cloud platforms, CDNs, etc.? There must be some “problems” behind.
  - The entire web is NOT “designed” to be like its current form; instead, it has been shaped by all these “solutions”, each was proposed because of certain “problems”

- In this course, we have discussed many “problems” related to website engineering, which have led to the need for rapid development tools, better maintainability & extensibility of code, better performance, … …, etc.
  - But in fact, these problems appeared almost since computers were invented decades ago
  - Solutions were also proposed and used well before the web was invented
    - E.g., standards/libraries/frameworks, “separation” of concerns, client/server architectures (or the “cloud” nowadays), standard API protocols, … …, etc.
  - The technologies we have introduced in this course are just **variants** of these **old** solutions, but tailored for the web
    - For example, the concept of MVC and SoC have been proposed and used since 1970s.
    - And as we will realize in the future, these same, old “problems” and “solutions” will appear again and again, but probably in different forms

- This course aims to promote understanding on website engineering through understanding these basic problems and solutions
  - With this understanding, one can easily master “new” technologies in the future because they are very likely to be just some **variants** of the old solutions to the **same, old problems**

- So, when we encounter a “new” web technology (which happens almost every month), we may ask: what existing technologies aim to solve the same problem?
  - Then most probably you would realize that they are conceptually similar. If not, a new problem has appeared (seldom happens though) which deserve further studies.

Examination

Objective and scope

- An open-book, online examination
  - => **Please refer to the Programme Office’s announcements (which will be issued in due course) for the final arrangements.**
- The exam will constitute **30%** of the course assessment
- The examination will be focused more on the conceptual understanding on the topics covered than the proficiency in programming
  - But a general understanding on the basic syntax of these programming elements is required
- Covered materials:
  - All powerpoint slides of lectures
  - But all slides and program code in labs and assignment will **NOT** be covered
  - All pre-/post-class readings, and all external links mentioned in lecture slides will **NOT** be covered

Tentative format (subject to revision)

- **A set of sample exam questions has been posted in Moodle as a reference**
  - Under the “Examination” section at the bottom of the Moodle page
- **The sample questions are taken from a past exam paper**
  - The format / style of some of the real exam questions may be similar to that of the sample questions.
- 70-80% short questions
  - Mainly analytical, e.g.,
    - Given some arbitrary real-life scenarios/website structure, how to improve them using the techniques introduced in the course?
    - What are the pros/cons/usage of a certain technology/standard?
    - How can a website be structured as an MVC application? How can that be extended to accommodate new functions/content?
    - How can a given web page be modified so that it can be loaded faster in a browser?
    - What are the pros and cons of the different web API protocols? How to model a RESTful/SOAP API?
- 20-30% programming
  - Simple; related hints/references may be provided whenever appropriate.
  - Grading will be focused on your understanding on the logic and flow of the program, but not only on their correctness
    - => If you don’t know some programming details, please provide the pseudo-code (some points would be deducted though)

Consultation

- Email (anytime - before and after exam); or
- Post to Moodle's forum (before exam)
- Consultation sessions
  - Small-group discussions on anything about web development are welcome – can be done online through Zoom
  - March 6, 2021

Reminders

- Assignment 1 – due on March 17, 2021

- About the group project:
  - Please send a project plan (i.e., the topic of your website, technologies and web APIs used, job distribution, etc.) to Steven by March 6, 2021 for approval.
  - Steven has posted a list of groups (and their members) to Moodle (under "News and Announcements"). If our record is not correct, please email us asap.

- Remember to submit the group project on time!
  - Deadline is April 10, 2021
  - No submission would be accepted after April 30, 2021

- If you have any questions, feel free to post to Moodle or email Steven or me.

Questions?

Optional consultation session

March 6, 2021, 2pm-5pm

via Zoom, or @P6-03, Graduate House

Optional consultation session
March 6, 2021, 2pm-5pm

- Will be conducted at P6-03, Graduate House. Steven and I will be there.

- **Optional** – for those who have questions on the labs, the assignment/project or any other course materials.

- You may come in person, **or** join our online Zoom meeting
  - For the online Zoom meeting, you may share your screen with us, or grant us the remote access (through Zoom) if you want us to look into the problem you have encountered.
  - We will discuss with the students one at a time, please be patient while you are waiting at the virtual "waiting room".

- Please read the announcement (at Moodle) for the Zoom link and other details.

Course evaluation

Course evaluation

- The online course evaluation system has not been opened yet, which will be opened later in the semester.
- You will receive an email from the Programme Office in due course.
- You are invited and welcome to provide feedbacks and comments on this course. Please check email.

Thank you!