A bit more about
Sorting
The Rise of Google

Announcement

- Merge sort game: how many people are participating?
- HW 3 to be posted later, deadline to be adjusted

Merge Sort And Quicksort The power of divide and conquer

Review: sorting

- Sorting algorithm: takes an array of data and puts it in (ascending or descending order)
- Making it easier to search (e.g. phone book)
- Many sorting algorithms: insertion sort, selection sort, merge sort, etc.

Today: divide and conquer sorting

Suppose you have an array of items to sort.

Divide and conquer:

1. Split the array into smaller arrays
2. Sort those smaller arrays (in the same way)
3. Combine the sorted small arrays to give the sorted large array
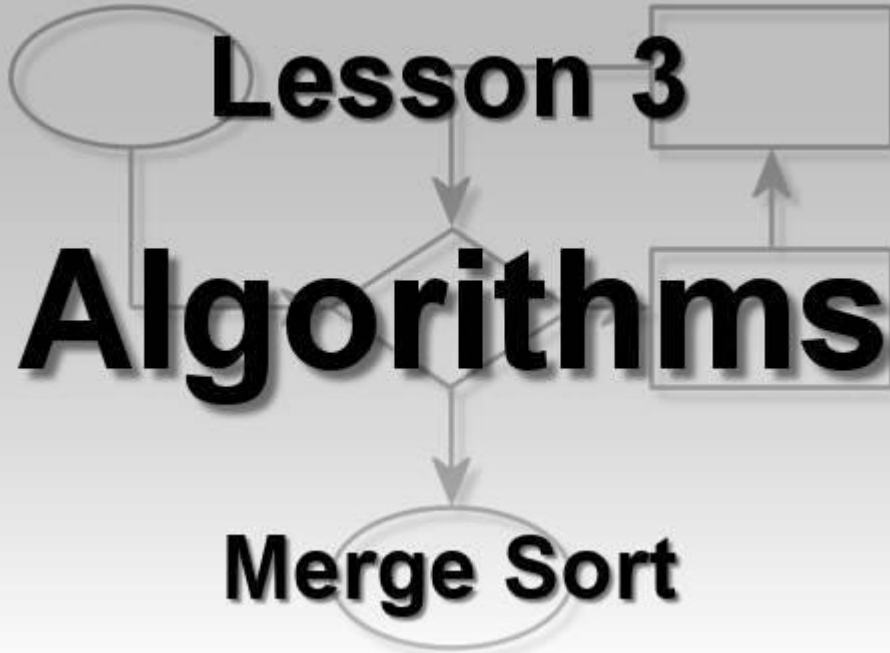
2 approaches

1. Easy split, hard to combine (Merge Sort)
2. Hard to split, easy to combine (Quicksort)

Merge Sort Game

https://www.youtube.com/watch?v=EeQ8pwjQxTM

https://www.youtube.com/watch?v=GCae1WNvnZM

Merge Sort Sketch

1. Divide the list into the smallest unit (1 element)
2. Compare each element with the adjacent list to sort and merge the two adjacent lists
3. All the elements are sorted and merged

Game Rule

1. Need 8 Volunteers to demonstrate merge sort
2. Bonus: 1 point each
3. Goal: learn divide and conquer sorting algorithm
4. 16 cups with numbers on both side
5. The 1st player divide the 16 cups into 8 sorted group, each with 2 cups
6. Each following player is going to sort 4, 8 or 16 cups using merge sort strategy
7. Until the entire set of cups are sorted: and the class pick out the errors during the sorting process

Quicksort

http://me.dt.in.th/page/Quicksort/

Quicksort

1. partition the array into two parts around a pivot
2. quicksort those smaller arrays
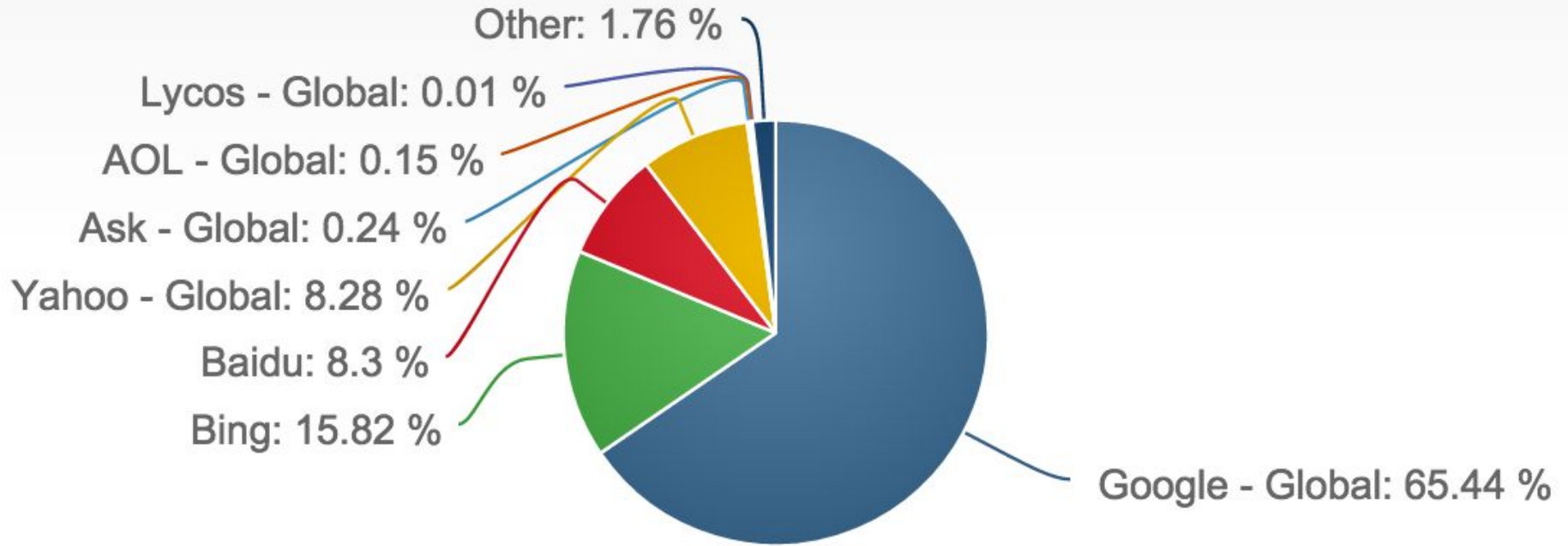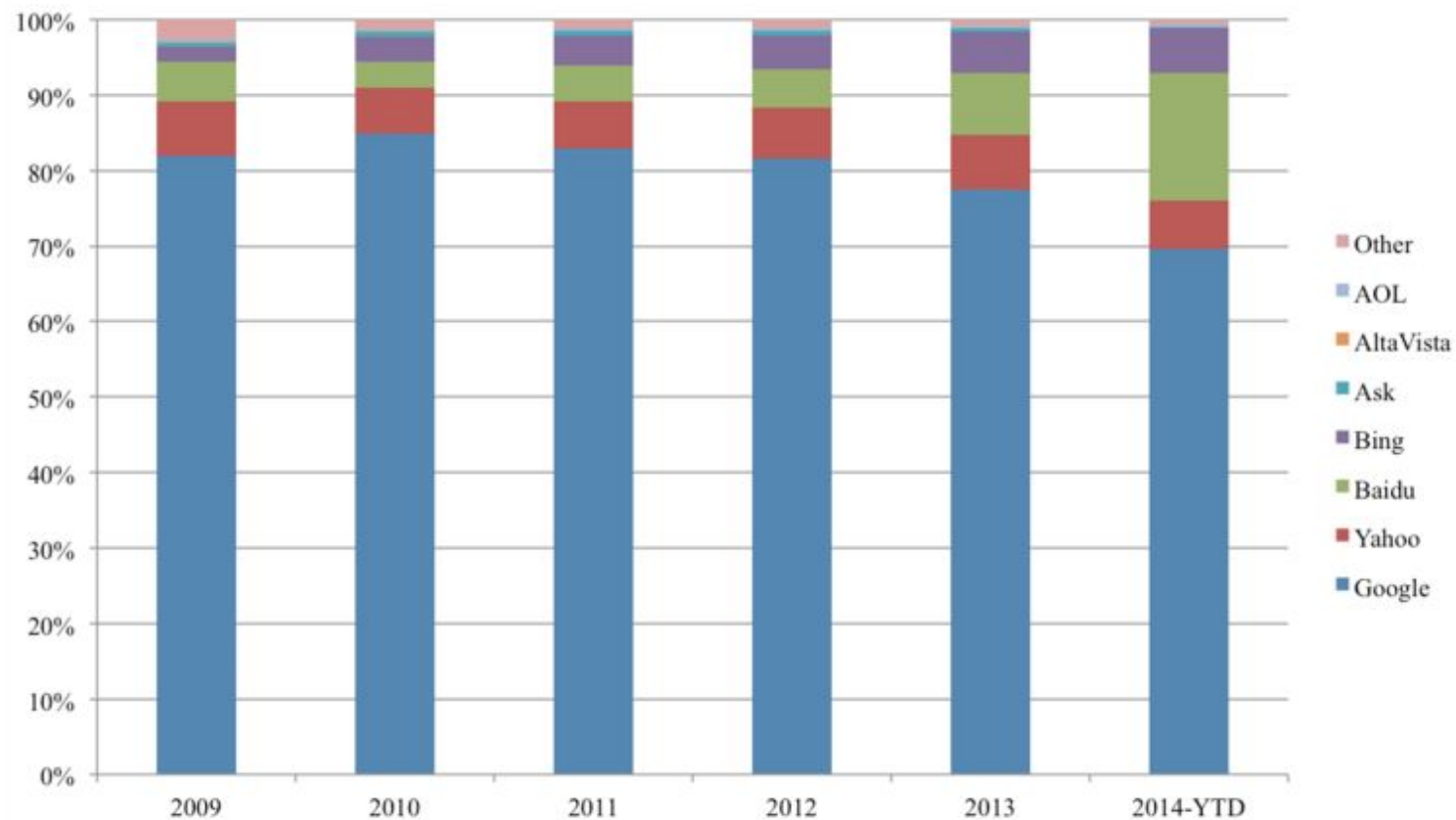3. concatenate the two sorted arrays end to end

Demo with cups

Ben Gomes
Google Fellow

http://www.google.com/insidesearch/howsearchworks/algorithms.html

SEARCH WARS

Desktop Search Engine Market Share, Jan 2016

Other: 1.76 %
Lycos - Global: 0.01 %
AOL - Global: 0.15 %
Ask - Global: 0.24 %
Yahoo - Global: 8.28 %
Baidu: 8.3 %
Bing: 15.82 %
Google - Global: 65.44 %

https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0

This chart shows the percentage of search engine market share on desktop, measured at six points in the last six years. Source: http://www.netmarketshare.com.

Why Google Succeeded (that is, becoming the world's most popular search engine) while others failed?

Some (arguably) key factors

- ☐ PageRank
- ☐ Better search results than competitors
- ☐ "a stubborn refusal to accept the orthodox view at the time, that stickiness was crucial to a website's success."
- ☐ Tried to sell tech to Excite: "Google was too good. If Excite were to host a search engine that instantly gave people information they sought, [Excite's CEO] explained, the users would leave the site instantly."
- ☐ Response to Excite: they do not know what they are talking about…

http://www.businessinsider.com/accurate-contrarian-theories-2011-5

Contrarian Investment Strategy

- **Contrarian**: opposing or rejecting popular opinion; going against current practice.
- Contrarian Investment Strategy: Going against the crowd
- "World-changing startups need to be premised on accurate contrarian theories"-- Investor Reid Hoffman

Google

- **Non-contrarian**: users would prefer a better search engine.
- **Contrarian**: it made business sense to get users off their site as quickly as possible -- against stickiness

http://www.businessinsider.com/accurate-contrarian-theories-2011-6

Simple interface: no annoying ads flooding, fastest to load and search and leave: winnin the users

More...

- Replaced bookmarks
- AdWords: 99% of revenue: Smart and Relevant Ads

http://www.bloomberg.com/bw/stories/2006-03-05/the-secret-to-googles-success
More readings:
The Google Story
Book by David A. Vise and Mark Malseed

And More…

- Technology: PageRank, the world's largest distributed computer system
- Business model innovation: targeted ads, better user experience –– highly effective revenue generator
- Brand: the most recognized brand in the world…it is a verb!
- Optimize user experience first, then revenue second

More readings:
http://www.ventureblog.com/2003/05/4-keys-to-googles-success.html
http://www.etiole.com/2008/12/5-reasons-why-google-did-succeed-but-others-did-not-must-read-for-googlers/

How web search engines work?
The Short Story

In near real time…

1. Crawling: when content is discovered
2. Indexing: when content is analyzed and stored in huge databases
3. Searching/Retrieval: a user query fetches a list of relevant pages

Crawling

- An automated bot – a spider/crawler – visits each page quickly.
- Scanning the site and getting a complete list of everything: page title, images, keywords, and other pages it links to, etc.
- Crawlers may cache a copy of the whole page, as well as page layout, where the advertising units are, etc.
- Crawler adds all the new links it found to a list of places to crawl next – in addition to re-crawling sites again to see if anything has changed. It's a never-ending process.
- Deep/Dark web: part of the WWW that's not indexed by the search engines

INDEXING

Taking all of that data you have from a crawl and placing it in a big database.

Google data center:

Ranking and Retrieval

- Type in a search query, google displays the most relevant documents it finds that match your query.
- Ranking algorithm: closely guarded secret
- Competition, prevent gaming the system
- Some ideas: keywords (stuffing), important based linking PageRank
- Focus right now: better content, user experience

Search Engine Optimization (SEO) gone wild!

find the people. Halloween.com and the Halloween.com list of Halloween sites will now help you find everything from Halloween history, Halloween greeting cards, Halloween screensavers, Halloween jokes, costumes for kids, costumes for adults, costumes ideas for the physically challenged, Halloween Movies, last minute costume ideas. From Halloween games, to Halloween safety to trick or treating information to Halloween events, and everything in between. Our Halloween discussion in the Halloween.com forums lets you exchange ideas, comments and fun about our favorite holiday. Halloween.com is ghouly your one source of all things Halloween! Our complete sitemap shows all the Halloween sections.
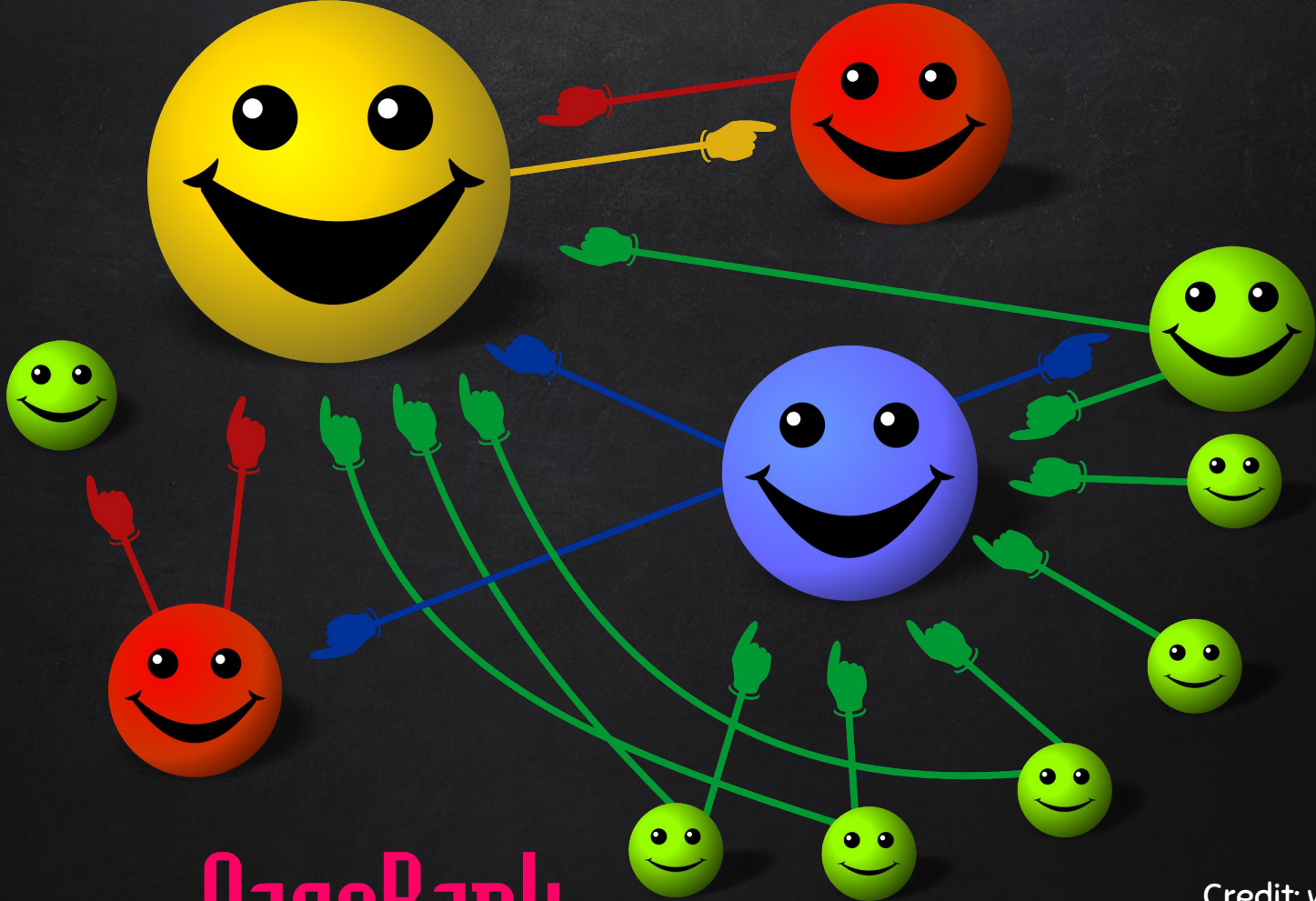
Happy Halloween!

PageRank
The Basics

PageRank

What is PageRank?

☐ How Google determines a page's relevance or importance.
☐ "PageRank" or "PR": a term to indicate the popularity of a page.
☐ The PR is determined by the number of links from other pages on the World Wide Web that point to this page.
☐ PR is like a vote by other pages in terms of its importance
☐ More votes, more important
☐ PR of the voters are also important in the computation
☐ Higher PR of voter page means better PR for the voted page

PageRank in Google

- PR does not directly influence a web page's ranking in the search engine results.
- PR doesn't determine which webpages are included in the search results when a search term is entered
- The search results ranking is determined by the relevance of titles, keywords and phrases contained within those pages.
- When two web pages have the same relevance to a search term, PR will determine which page is displayed first in the search results.
- PR is very important for **search engine optimization** (SEO)

THANKS!

Any questions?

You can find me at
beiwang@sci.utah.edu

http://www.sci.utah.edu/~beiwang/teaching/cs1060.html

CREDITS

Special thanks to all the people who made and released these awesome resources for free:

- ☐ Presentation template by SlidesCarnival
- ☐ Photographs by Unsplash