A Closer Look at Computing: The Million-Dollar Algorithms

PHILIP CHAN
RYAN STANSIFER

MAR 24, 2017

Ranking Web Pages

PHILIP CHAN

Basic Problem

- Input:
  - Search query (keywords)
  - Web pages
- Output
  - Ranking of web pages relevant to the search query

Basic Problem

- Input:
  - Search query (keywords)
  - Web pages
- Output
  - Ranking of web pages relevant to the search query
- If given 10 web pages and a search query (e.g. internet)
  - How would your rank the web pages?

Search Engines in the 1990s

- Altavista
- Excite
- Infoseek
- Webcrawler
- …

- Internet Archive: archive.org/web

Search Engines in the Early 2000s

- Many search engines faded away
- Google began to dominate

Secret Sauce of Google

- Google did not advertise itself
  - Mostly by word of mouth
  - “free advertisement” from the press
- Why did Google start to dominate?
- What is the secret sauce of Google?

Brief History of Google

- 1996 -- 2 graduate students at Stanford U.
- 1997 – google.com was registered
- 1998 – [to be filled in later]
- 1998 – $100K investment, company incorporated
- 2004 – initial public offering (NASDAQ)

Secret Sauce of Google

- A different way to rank web pages
- In addition to words on a page
  - They use links between pages
- How would you use links to rank pages?

Brief History of Google

- 1996 -- 2 graduate students at Stanford U.
- 1997 – google.com was registered
- 1998 -- publication of their algorithm in WWW Conf.
- 1998 – $100K investment, company incorporated
- 2004 – initial public offering (NASDAQ)

Algorithm: Page Rank

- Input
  - Links between web pages
- Output
  - Score for each web page
- https://www.youtube.com/watch?v=Q9HjeFD62Uk
  - 8:27-13:43

Mystery Million-Dollar Problem

PHILIP CHAN

What is common among these tasks?

- With a grocery shopping list
  - you try to gather all the items and leave quickly
- Given a truck load of packages
  - a UPS/FedEx driver tries to deliver them quickly
- Given locations of holes on a circuit board
  - Drill all the holes quickly

Common Problem

- Given an initial location and other locations
  - Find the shortest route:
    - Start from an initial location
    - Visit each of the other locations exactly once
    - Go back to the initial location
- Traveling Salesman Problem (TSP)

Traveling Salesman Problem (TSP)

- https://www.youtube.com/watch?v=Q9HjeFD62Uk

- 36:12- 38:50

A Closer Look at TSP

- Consider the initial location is Home
  - A, B, C are other locations

	1st visit	2nd visit	3rd visit	
Home	A			
	B			
	C			

A Closer Look at TSP

- Consider the initial location is Home
  - A, B, C are other locations

	1st visit	2nd visit	3rd visit	
Home	A	B		
		C		
	B	A		
		C		
	C	A		
		B		

A Closer Look at TSP

- Consider the initial location is Home
  - A, B, C are other locations

	1st visit	2nd visit	3rd visit	
Home	A	B	C	
		C	B	
	B	A	C	
		C	A	
	C	A	B	
		B	A	

A Closer Look at TSP

- Consider the initial location is Home
  - A, B, C are other locations

	1st visit	2nd visit	3rd visit	Last visit
Home	A	B	C	Home
		C	B	
	B	A	C	
		C	A	
	C	A	B	
		B	A	

A Closer Look at TSP

- Consider the initial location is Home
  - A, B, C are other locations

	1st visit	2nd visit	3rd visit	Last visit
Home	A	B	C	Home
		C	B	
	B	A	C	
		C	A	
	C	A	B	
		B	A	

  - 3 * 2 * 1 = 6 possible routes

4 locations

- 1st visit
  - 4 choices
- 2nd visit
  - 3 choices
- 3rd visit
  - 2 choices
- 4th visit
  - 1 choice
- 4*3*2*1 = 24 possible routes

n locations

- 1st visit
  - n choices
- 2nd visit
  - n-1 choices
- …
- nth visit
  - 1 choice
- # of possible routes?

n locations

- 1st visit
  - n choices
- 2nd visit
  - n-1 choices
- …
- nth visit
  - 1 choice
- # of possible routes?
  - n*(n-1)*(n-2) …. *2*1 = n!

Computer checks 1 billion routes per second

# of locations	# of possible routes	Completion time
10	3.6M	0.003 s

Computer checks 1 billion routes per second

# of locations	# of possible routes	Completion time
10	3.6M	0.003 s
15		?

Computer checks 1 billion routes per second

# of locations	# of possible routes	Completion time
10	3.6M	0.003 s
15	1.3 x 10^12	22 minutes
20		?

Computer checks 1 billion routes per second

# of locations	# of possible routes	Completion time
10	3.6M	0.003 s
15	1.3 x 10^12	22 minutes
20	2.4 x 10^18	77 years
100		?

Computer checks 1 billion routes per second

# of locations	# of possible routes	Completion time
10	3.6M	0.003 s
15	1.3 x 10^12	22 minutes
20	2.4 x 10^18	77 years
100	9.3 x 10^157	3 x 10^141 years

TSP and Bees

- https://www.youtube.com/watch?v=Q9HjeFD62Uk

- 38:50-42:30

$1M prize

- Clay Mathematical Institute
  - claymath.org
- Millennium Prize Problems (7)
  - $1M each
- P=NP?
  - Does an efficient algorithm for TSP exist?
    - A yes or no answer will win the prize
    - If yes, (a large number of) problems in the same class can be solved efficiently.

Part 2

Thank You

- cs.fit.edu/~pkc/cs4hs
  - Summer Camps
    - July
- 8pm, astronomy lecture, Olin Engineering, Rm 118
  - Across the courtyard
  - Between this building and the parking lot
- 9pm, 32-inch telescope
- Questions?