122COM: Searching

David Croft

Introduction

Linear search

Rinary spare

String searchin

Recar

122COM: Searching

David Croft

Coventry University david.croft@coventry.ac.uk

2015



Overview

- 1 Introduction
- 2 Linear search
- 3 Binary search
- 4 String searching
- 5 Recap



Introduction
Linear search
Binary search

Searching is used everywhere in computing.

- Obvious applications.
 - Text files.
 - Databases.
 - File systems.
- Hidden applications.
 - Computer games.
 - FOV search for objects in view.



- Path finding algorithms in games
 - https://www.youtube.com/watch?v=19h1g22hby8
- Brute force approaches that find the best/shortest/fastest problem are too slow (travelling salesman).
- Heuristic aproaches are used instead.
 - Find "good enough" solutions.
 - Not always the best solution.
 - Dijkstra's algorithm.
 - A* algorithm.



Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R
\uparrow	•													
Ζ														



ntroduction .

Linear search

String

кеса

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R	
	\uparrow			•			•								
	7														



Simplest search. Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.

			_		_		•		_				13		
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R	
		\uparrow		•	•		•			•					
		Ζ													



Linear search

Also called sequential search.

- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.

- O(n)
 - Will discuss *O*() notation in a later week.





Linear search

Also called sequential search.

Iterate over elements.

Simplest search.

Until found or until end of sequence.

■ Potentially slow.

O(n)

■ Will discuss *O*() notation in a later week.

													13	
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R





Introduction Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.



Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.

- O(n)
 - Will discuss *O*() notation in a later week.

														14
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R



Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.





ntroduction

Linear search

String

Reca

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.

															14
A	4	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R





Simplest search. Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.





Introduction Linear search

Also called sequential search.

Iterate over elements.

Until found or until end of sequence.

■ Potentially slow.

Simplest search.

O(n)

■ Will discuss *O*() notation in a later week.

													13	
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R



Introduction Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.

- O(n)
 - Will discuss *O*() notation in a later week.



ntroduction

Linear search

String

Reca

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.





Introduction Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.

- O(n)
 - Will discuss *O*() notation in a later week.

														14
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R







ntroduction

Linear search SIM

String

кеса

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.

														14
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R





Introduction Linear search

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.

- O(n)
 - Will discuss *O*() notation in a later week.

															14
A	4	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R





Simplest search. Linear search

Also called sequential search.

Iterate over elements.

Until found or until end of sequence.

■ Potentially slow.

O(n)

■ Will discuss *O*() notation in a later week.

															14
A	4	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R





ntroduction

Linear search

String

кеса

- Also called sequential search.
- Iterate over elements.
- Until found or until end of sequence.
- Potentially slow.
- O(n)
 - Will discuss *O*() notation in a later week.

														14
Α	В	Z	Q	K	L	G	Н	U	Α	Р	L	F	N	R





Muuuuuuch faster than linear search.

- Divide & conquer.
- Only works on sorted sequences.
- Algorithm is:
 - Find middle value of sequence.
 - If search value == middle value then success.
 - If search value is < middle value then forget about the top half of the sequence.
 - If search value is > middle value then forget about the bottom half of the sequence.
 - Repeat from step 1 until len(sequence) == 0.



				-			-							14
Α	В	C	D	Е	F	G	Н	I	J	K	L	M	Ν	0



Linear search
Binary search

String searchin

Pocan

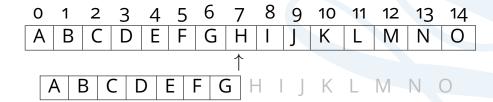


Introduction
Linear search

Binary searcl

searchin

Recap

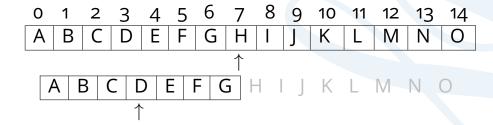




Linear search Binary search

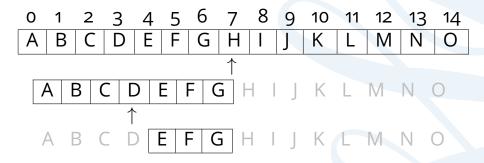
String searching

Recap





Binary search

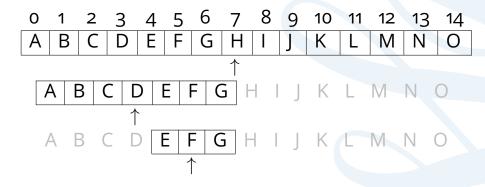




Linear search
Binary search

String searching

Recap

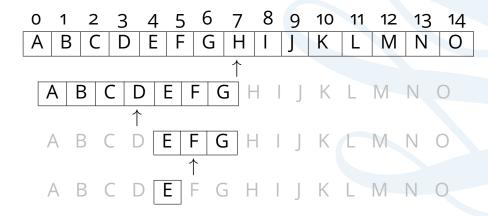




Linear search Binary search

String

Recap



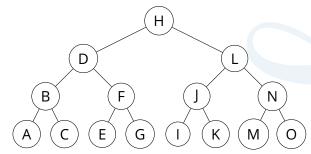


Binary search



How many comparisons do we need to do for binary search?

■ How many times can we divide our list by 2?

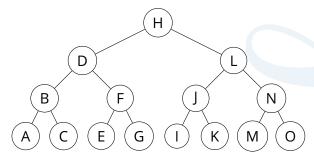






Binary search

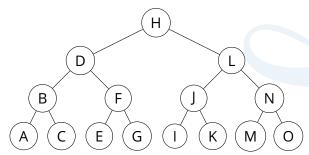
- How many times can we divide our list by 2?
- Ideally depth of tree is $log_2(n)$
 - n = 14.
 - $\log_2(14) = 3.9 \Rightarrow 3$



Í

Binary search

- How many times can we divide our list by 2?
- Ideally depth of tree is $log_2(n)$
 - n = 14.
 - $\log_2(14) = 3.9 \Rightarrow 3$
- Binary search has a complexity of $O(\log n)$.
 - Will cover *O*() complexity in later week.

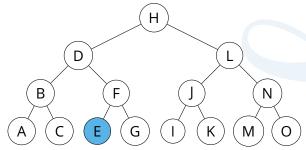




Introduction

Binary search

- How many times can we divide our list by 2?
- Ideally depth of tree is $\log_2(n)$
 - n = 14.
 - $\log_2(14) = 3.9 \Rightarrow 3$
- Binary search has a complexity of $O(\log n)$.
 - Will cover *O*() complexity in later week.
- Find E.





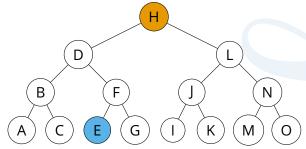
Introduction

Linear search

String searching

кеса

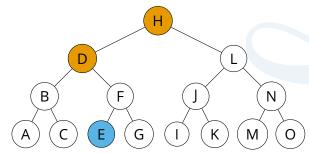
- How many times can we divide our list by 2?
- Ideally depth of tree is $log_2(n)$
 - n = 14.
 - $\log_2(14) = 3.9 \Rightarrow 3$
- Binary search has a complexity of $O(\log n)$.
 - Will cover *O*() complexity in later week.
- Find E.





Binary search

- How many times can we divide our list by 2?
- Ideally depth of tree is $log_2(n)$
 - n = 14.
 - $\log_2(14) = 3.9 \Rightarrow 3$
- Binary search has a complexity of $O(\log n)$.
 - Will cover *O*() complexity in later week.
- Find E.



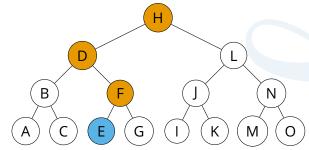


Linear search

String searching

How many comparisons do we need to do for binary search?

- How many times can we divide our list by 2?
- Ideally depth of tree is $log_2(n)$
 - n = 14.
 - $\log_2(14) = 3.9 \Rightarrow 3$
- Binary search has a complexity of $O(\log n)$.
 - Will cover *O*() complexity in later week.
- Find E.







Binary search

Clearly much faster than linear search.

- To search a trillion elements linearly could mean a trillion comparisons.
- 40 with binary search.

But...

- Have to sort the list first.
- Sorting lists can be expensive.
- Can't always sort sequences.
- Ordering is important.
- Cant always search for sequences.
 - Text documents.
 - Genetic codes.





Linear search

String searching

Recan

I.e. Text searching.

Finding one sequence in another sequence.

е

Х

а

- Naive search.
 - Like linear search.
 - Is very slow.

m p

e



◆□▶◆□▶◆■▶◆■▶ ■ り9



String searching

Boyer-Moore string searching algorithm.

_ 1077

1977.

■ Not going to talk about the whole algorithm here.

Gets really complex.

Right to left comparison.

Can skip sections of the text.

Don't need to test every position.

How?



String searching



Boyer-Moore string searching algorithm.

- **1977.**
- Not going to talk about the whole algorithm here.
 - Gets really complex.
- Right to left comparison.
- Can skip sections of the text.
 - Don't need to test every position.
- How?
- Pre-processes the search string.
 - Bad character rule table.
 - Explained in a minute.



example
$$\Rightarrow \frac{a \quad e \quad l \quad m \quad p \quad x \quad *}{4 \quad 6 \quad 1 \quad 3 \quad 2 \quad 5 \quad 7}$$

Linear search

Binary searc

String searching

Jearen



Linear search

Binary searc

String searching

Recap



searching



Recap



Linear search

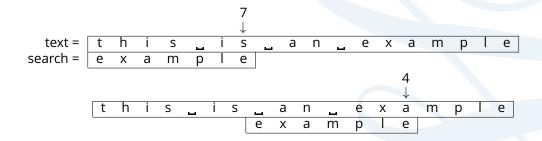
Binary searc

searching

Recan



searching

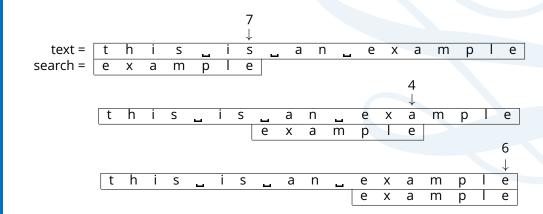




Linear search

String

searching





Linear search

String searching

Reca

- For each character.
- Just count number of places between it and end of search string.

example
$$\Rightarrow$$
 a e l m p x *



Linear search

String searching

Reca

- For each character.
- Just count number of places between it and end of search string.

example
$$\Rightarrow \frac{a \quad e \quad l \quad m \quad p \quad x \quad *}{4}$$



Reca

Boyer-Moore III

- For each character.
- Just count number of places between it and end of search string.

example
$$\Rightarrow \frac{a \quad e \quad l \quad m \quad p \quad x \quad *}{4 \quad 6}$$



Rinary sear

String searching

Reca

- For each character.
- Just count number of places between it and end of search string.

example
$$\Rightarrow \frac{a \quad e \quad l \quad m \quad p \quad x \quad *}{4 \quad 6 \quad 1}$$



- For each character.
- Just count number of places between it and end of search string.

example
$$\Rightarrow \frac{a \quad e \quad l \quad m \quad p \quad x \quad *}{4 \quad 6 \quad 1 \quad 3}$$



miroduction

Linear search

String searching

Reca

- For each character.
- Just count number of places between it and end of search string.



Reca

- For each character.
- Just count number of places between it and end of search string.



Reca

- For each character.
- Just count number of places between it and end of search string.



searching



Doesn't need to sort or modify the sequence being searched.

■ Small amount of pre-processing on the search value.

Worst case.

Linear time.

Average case

Sub-linear.

Not the only string searching algorithm.

- Knuth-Morris-Pratt.
- Finite State Machine (FSM).
- Rabin-Karp.





122COM: Searching

David Croft

Introduction

Linear search

Rinary sear

String searching

Recap

Quiz



- Searching
 - Applications everywhere.
 - Linear search.
 - Simple.
 - Slow.
 - Binary search.
 - Ordered sequence.
 - Very fast.
 - String searching.
 - Finding subsequence in sequence.
 - Boyers-Moore.
 - Preprocessing.
 - Skipping sections.



122COM: Searching

David Croft

Introduction

Linear search

Binary sear

String searchin

Recap

The End

