

# Prefix-querying Sequences Efficiently with Tries

---



**Rasmus Resen Amossen**

SOLUTION ARCHITECT

[rasmus.resen.org](http://rasmus.resen.org)

# The Match Finder App

Hashing

My Pictures

Hash functions

Exact tables

Probabilistic filters



Spatial index trees

Disjoint-Set structures



Cooking

Guitar



Tries

ing

Suffix trees

Cooking

Fishing

Ingredients

King Kong

Kings

# Autocomplete

From Wikipedia

data

Q

**Data**

**Database**

**Data** mining

**Data** center

**Data** compression

**Data** analysis

**Data** model

**Data** warehouse

**Data** rate units

**Data** Encryption Standard

containing...

*data*

Demo

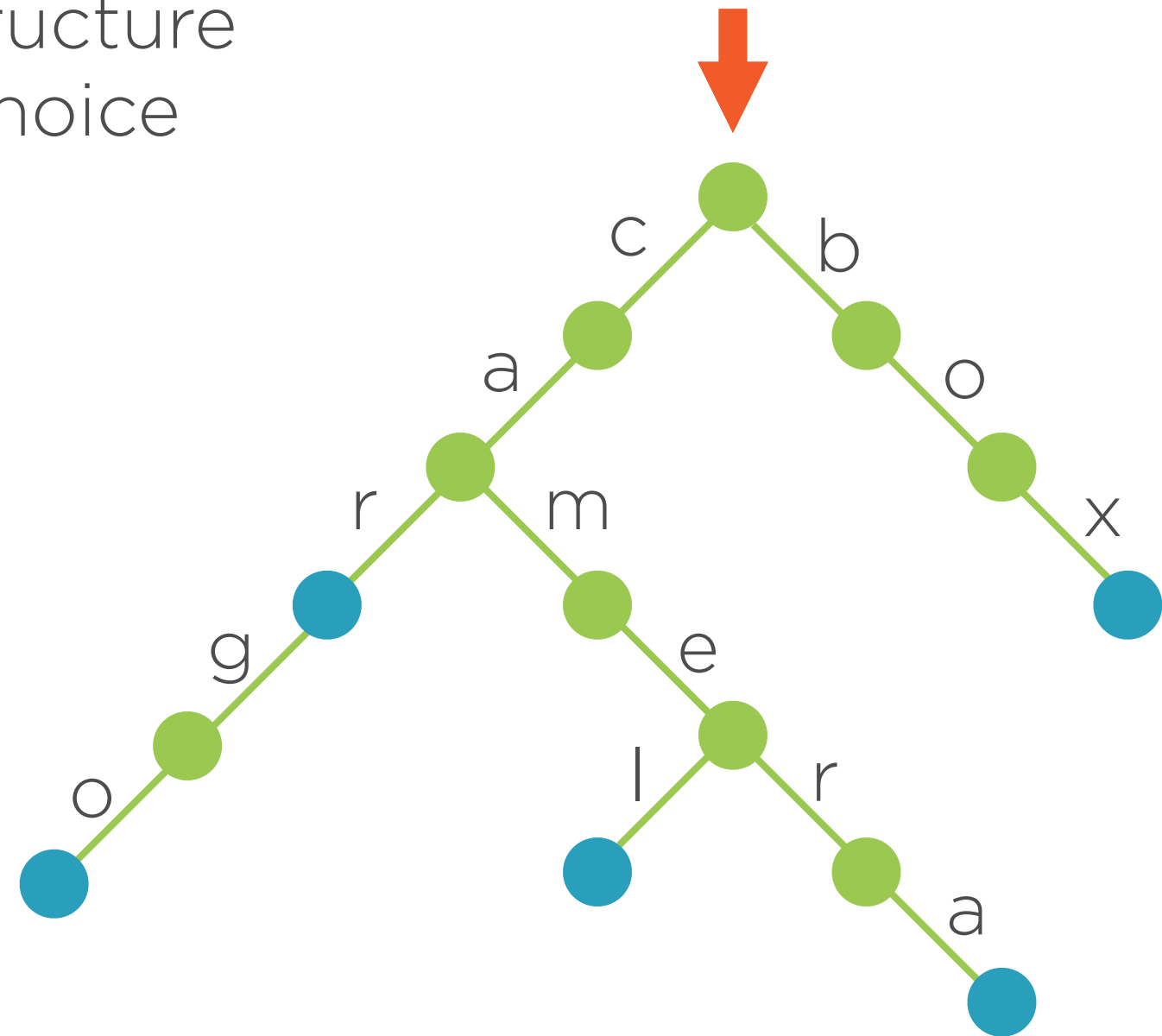
**Beginning an Autocompletion**

# Tries

Represent strings in a single tree structure  
Each edge represents a character choice

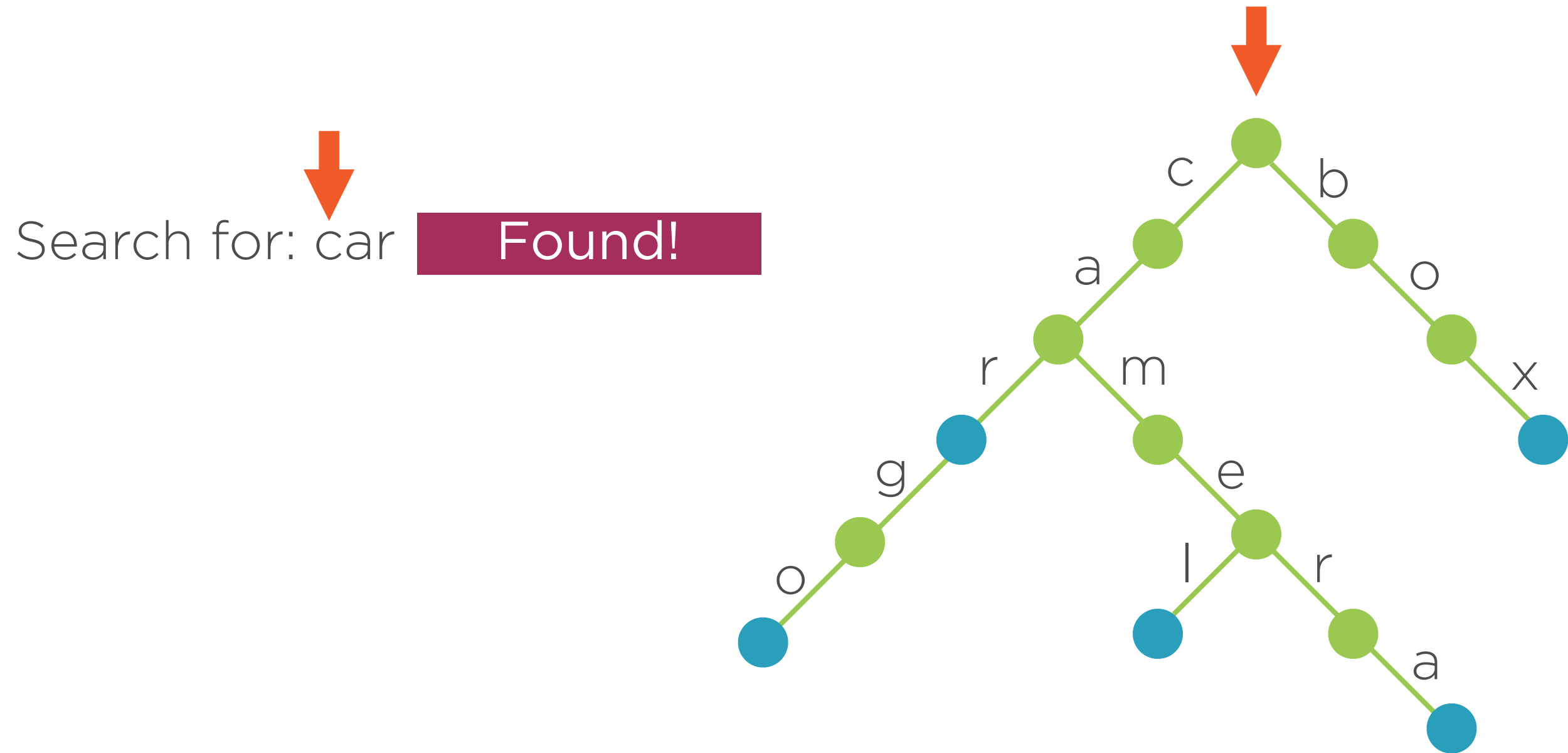
## Node

- isWord: Boolean
- Children: Hash map character  $\rightarrow$  node



Construction time:  $\Theta(N)$  for  $N$  characters in total  
Space usage:  $O(N)$

# Querying



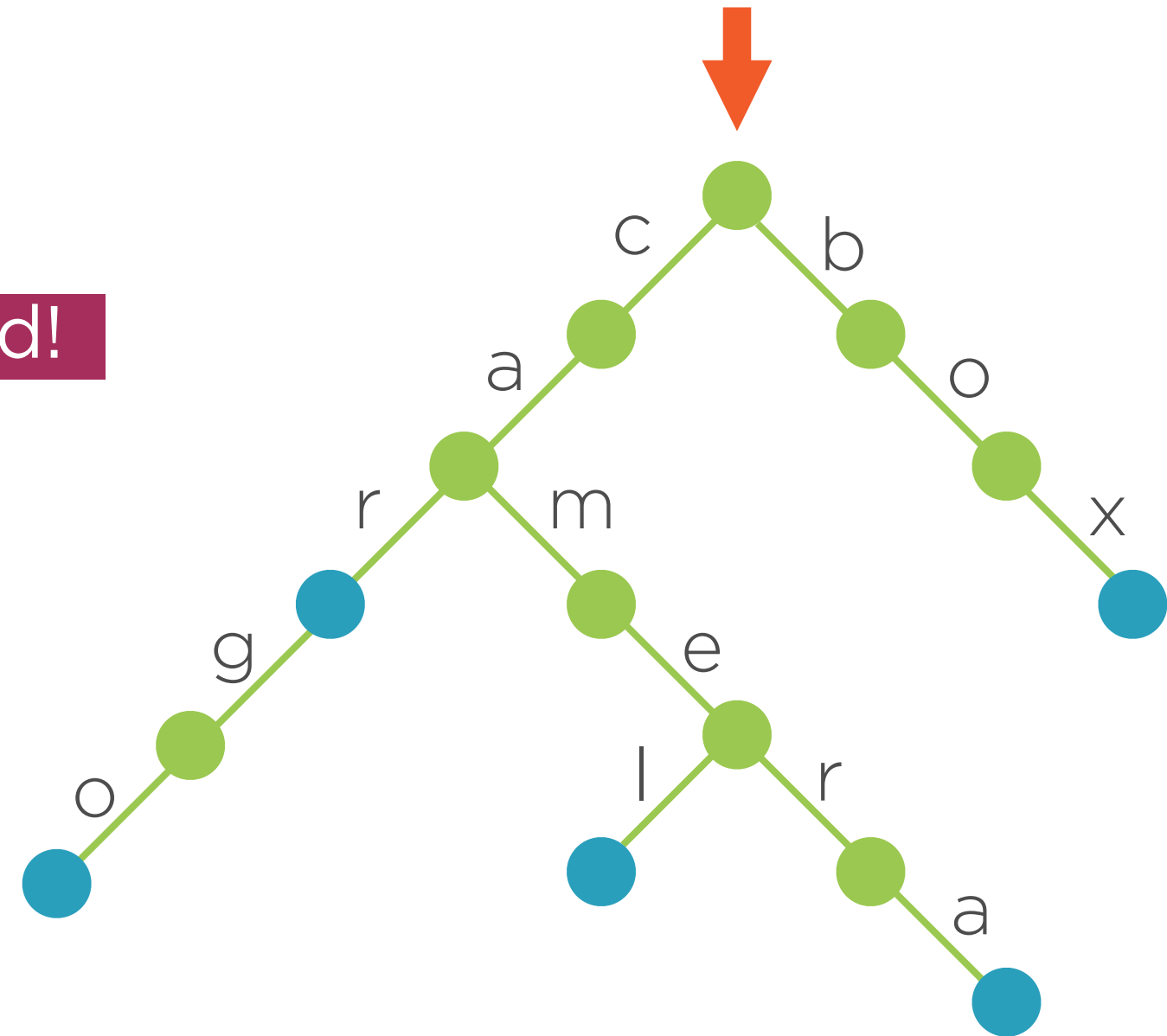
Complexity:  $O(N)$  for query size of  $N$



# Querying

Search for: cam

Not found!

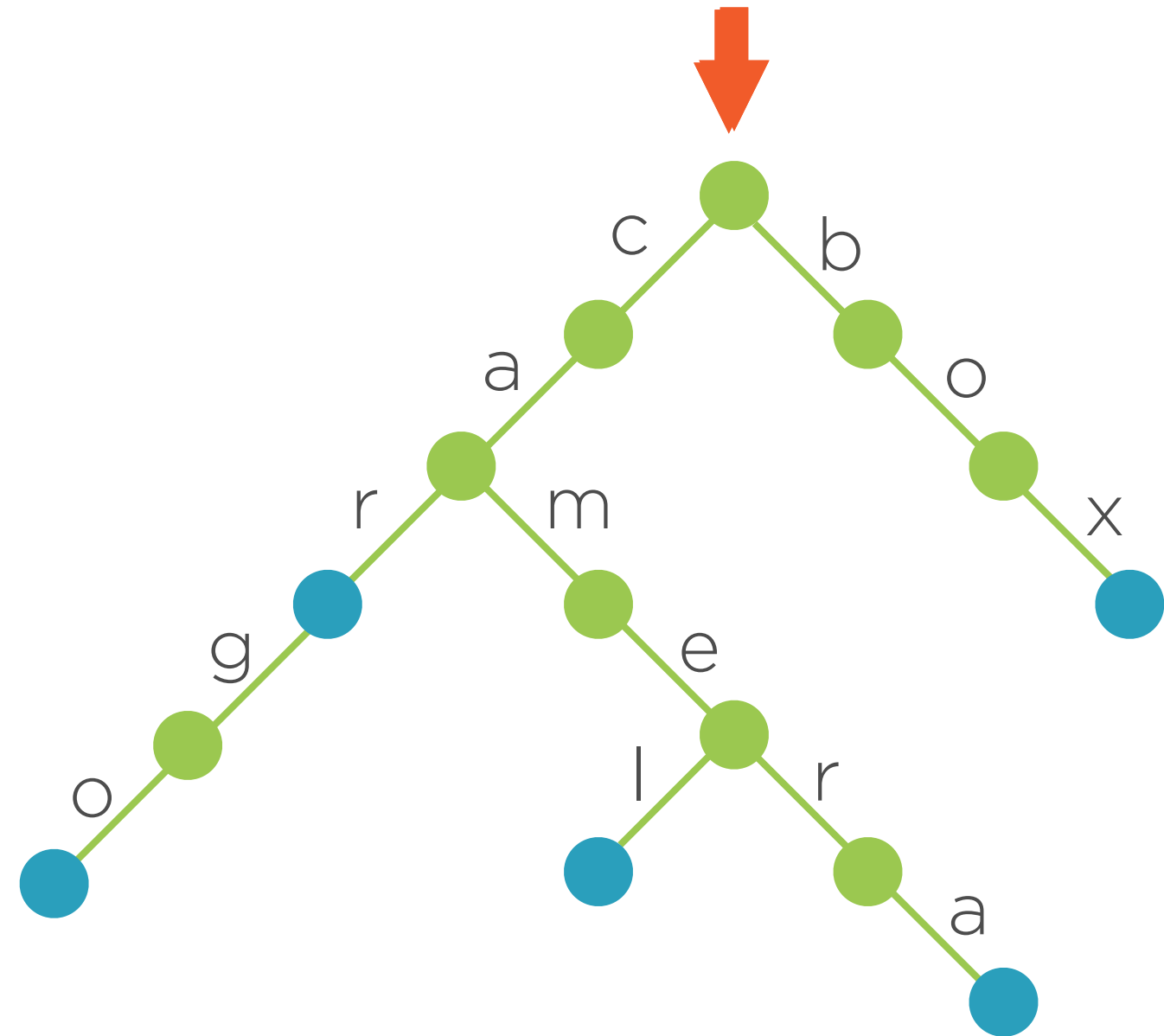


# Iterative Querying

Search for: car

Always starting from root:  
 $1 + 2 + 3 + \dots + N$  node visits  
That is  $O(N^2)$  in total

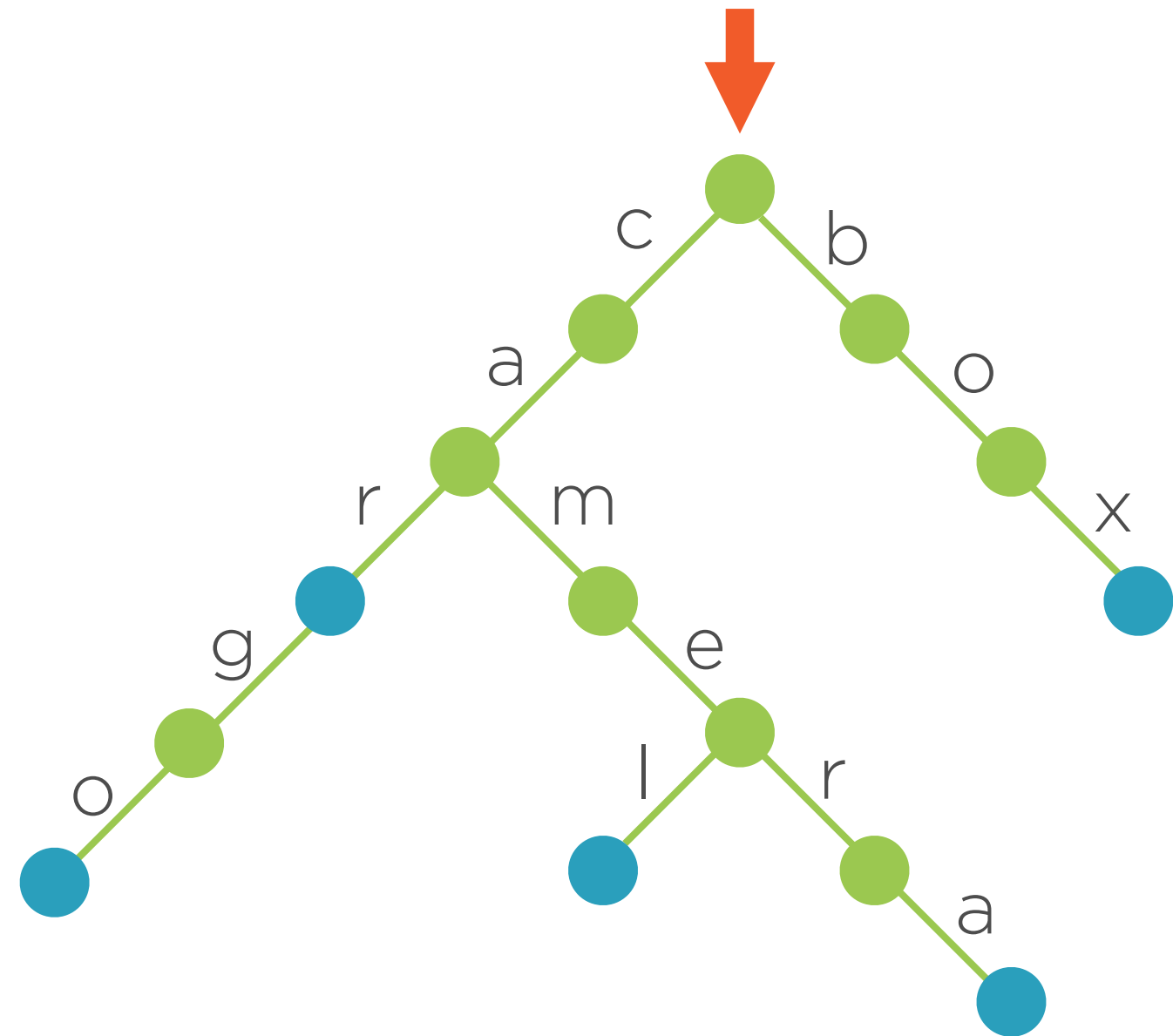
Save state:  
 $1 + 1 + 1 + \dots + 1$  node visits  
That is  $O(N)$  in total





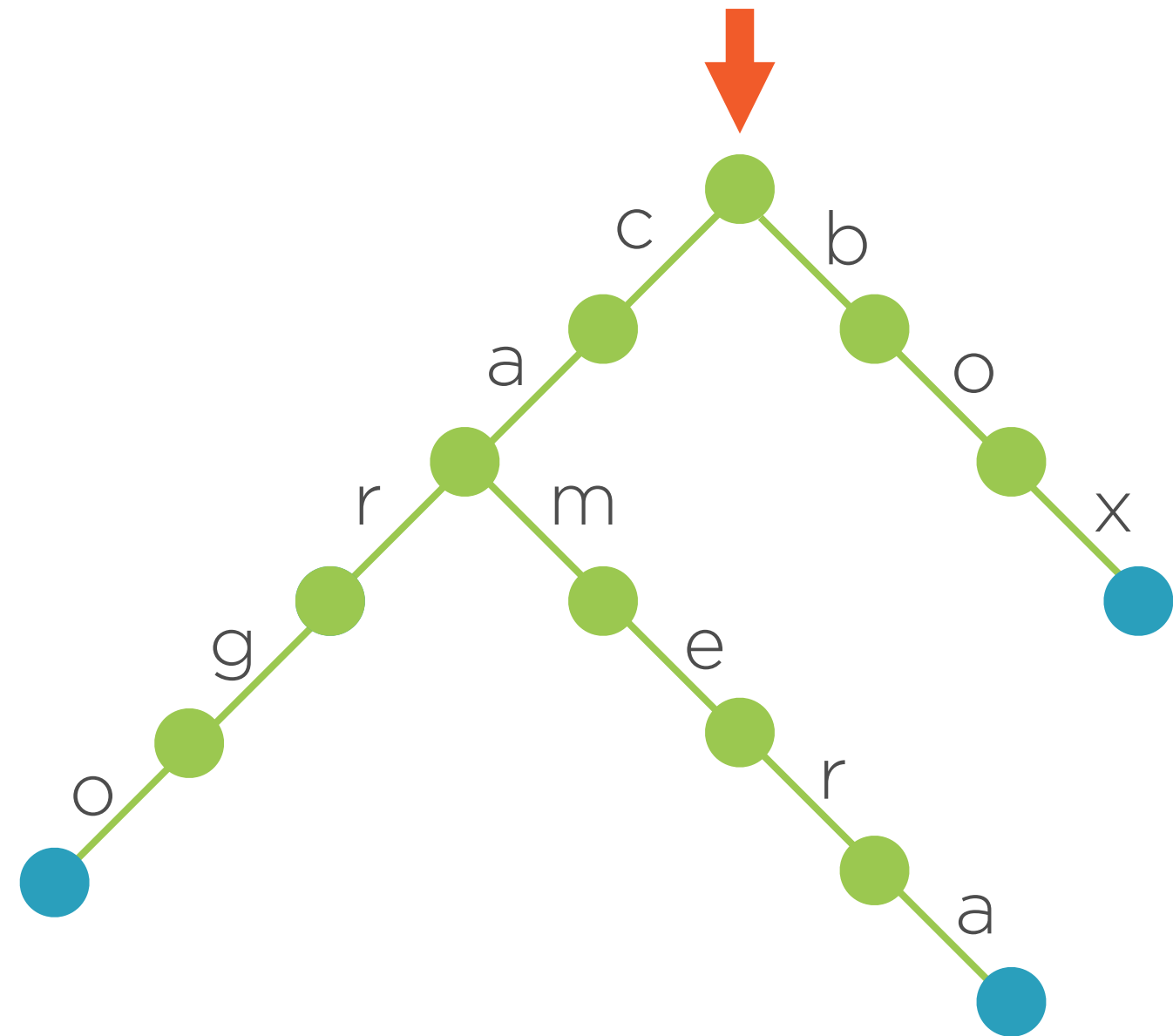
# Deletion

Delete: camel



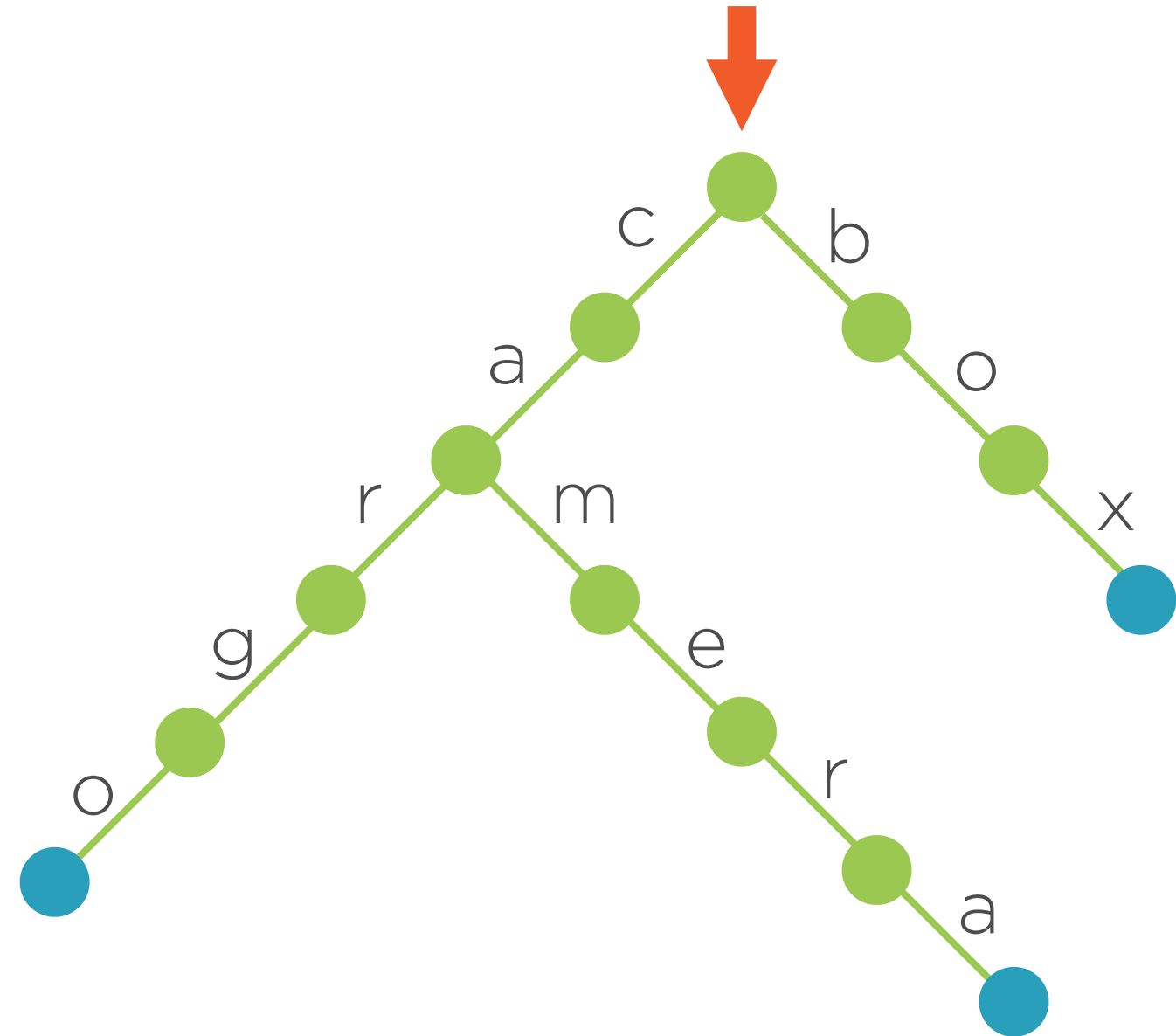
# Deletion

Delete: car



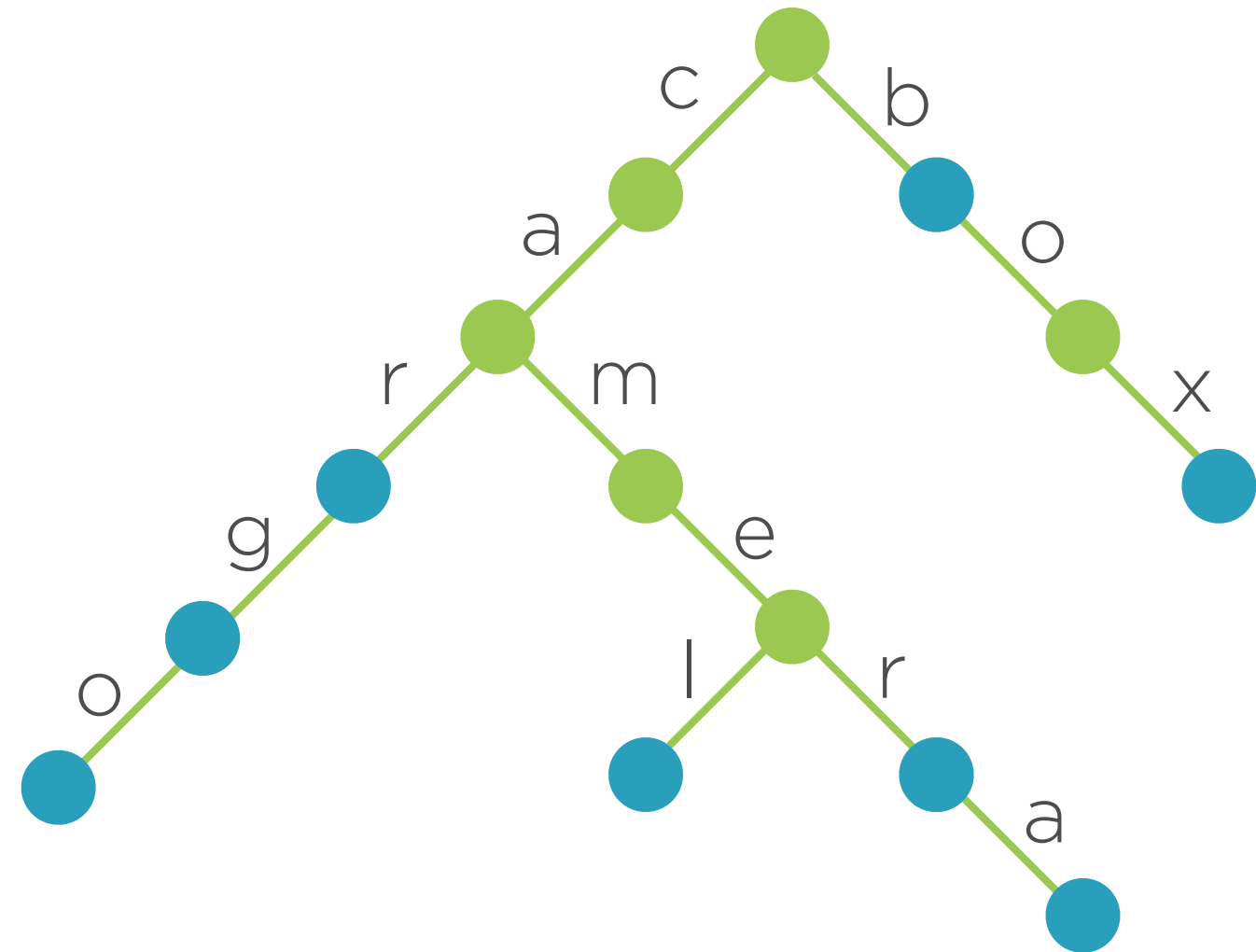
# Deletion

Delete: cargo



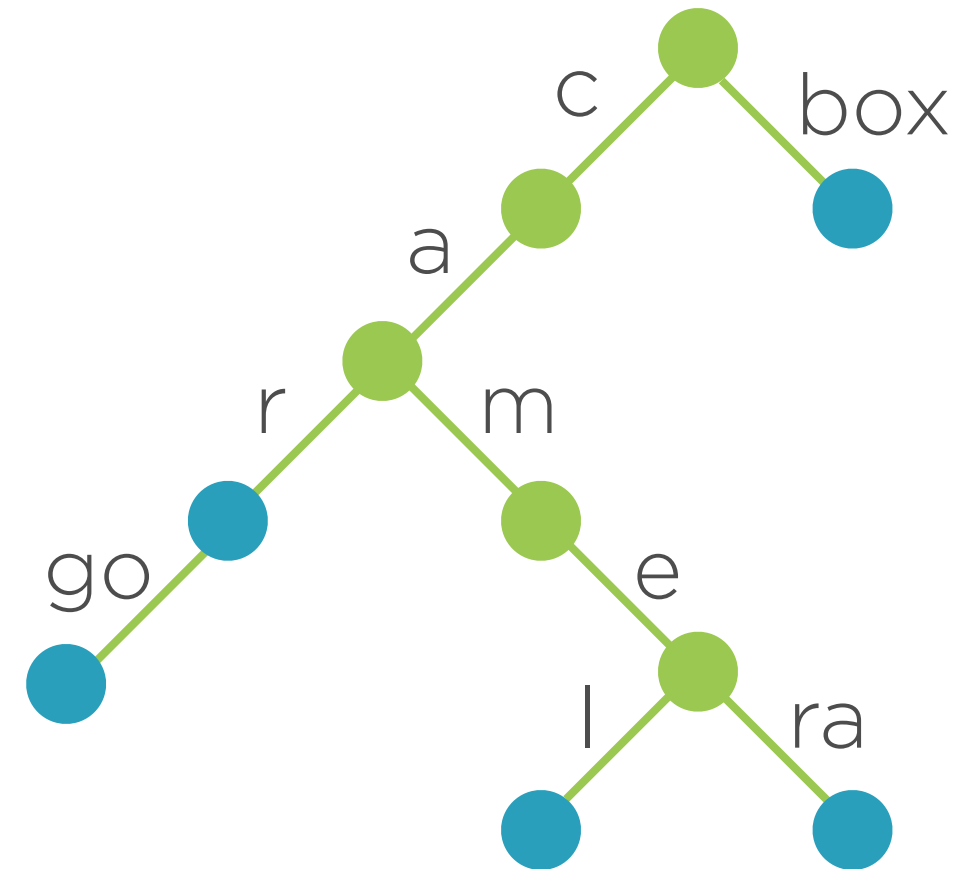
Complexity:  $O(N)$  for a word with  $N$  letters

# Radix Tree (compressed trie)



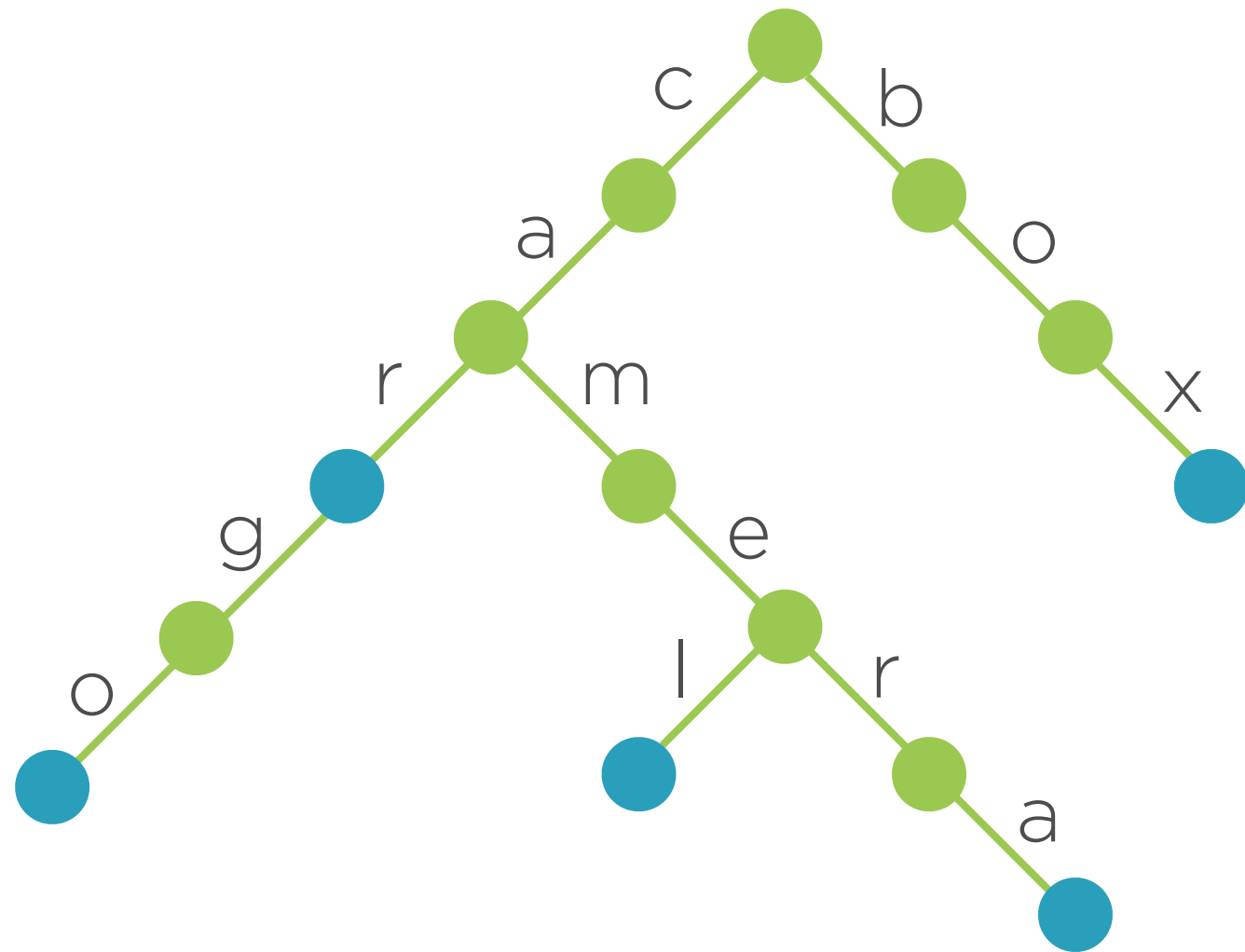
Minimum number  
of children

# Radix Tree (a compressed trie)

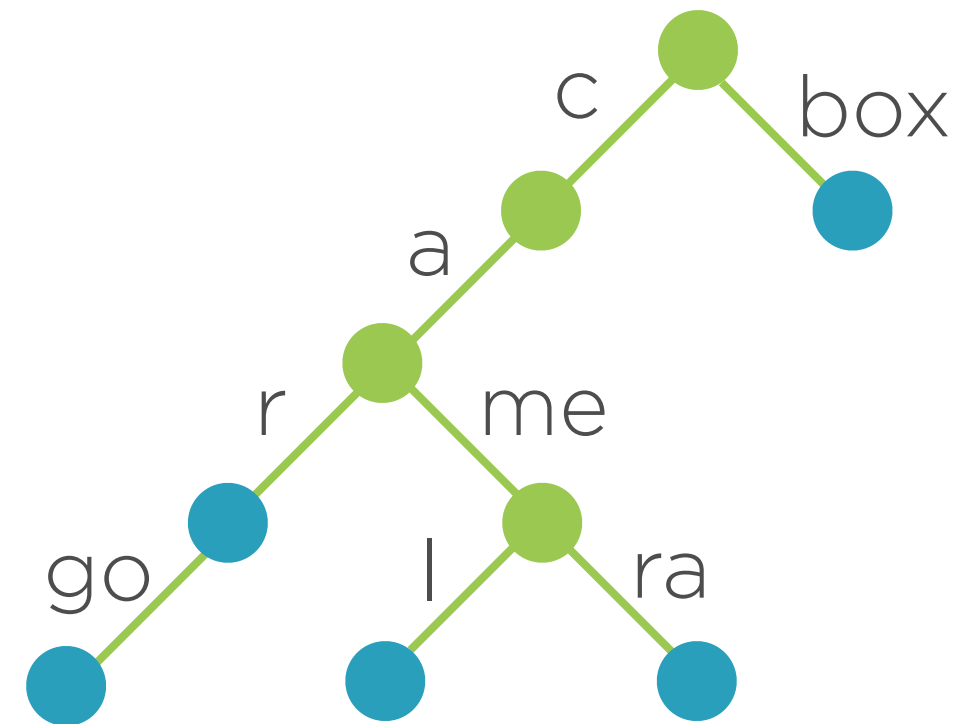


Minimum number  
of children

# Radix Tree (a compressed trie)



[tinyurl.com/triePlayground](http://tinyurl.com/triePlayground)  
[tinyurl.com/radixPlayground](http://tinyurl.com/radixPlayground)



Radix = 2:  
Patricia Tree

Useful if:  
Initialize once  
Then querying

Demo

**Implementing the Autocompletion**



# Lessons Learned

Useful for autocompletion

Prefix only

Query time

proportional with query string

independent on total amount of text

Consider compression if read-only  
(or read-mostly)