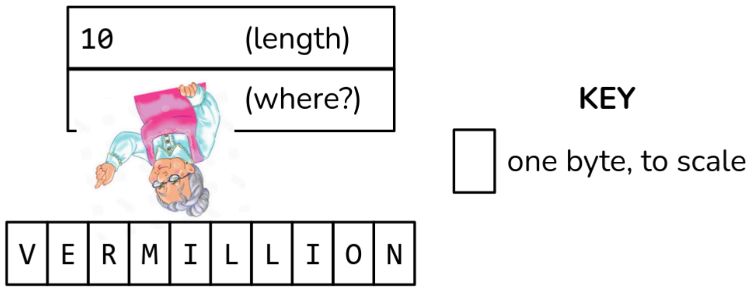
Wade Cheng | wadec@andrew.cmu.edu
https://github.com/wade-cheng/sso-talk

Small String Optimization

What are strings/why optimize?

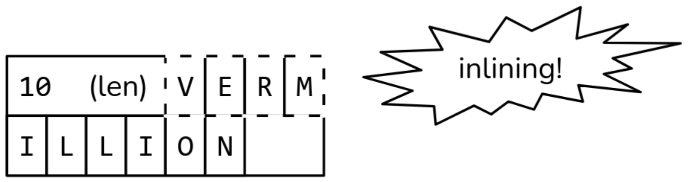
- strings: of characters!
    - websites you browse
    - name and passwords for bank transactions
    - character played in game
    - etc etc...
- modern databases store LOTS of data; we need the performance!

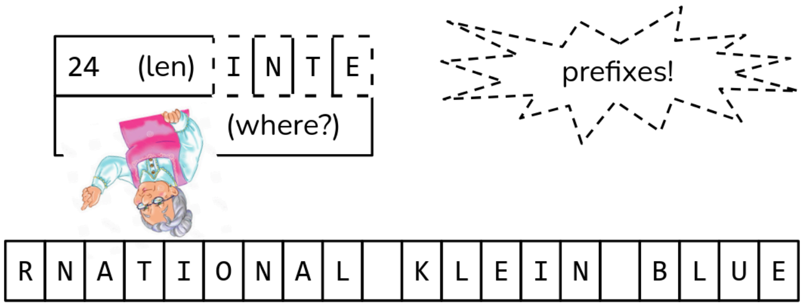
Technical overview of strings



- following the pointer grandma is slow!

How we can optimize





- inlining: in short strings, put the string where the pointer is. we never use the pointer grandma!
- prefixes: operations like string comparison don't need to follow pointer grandma until the 5th letter

Further Reading

https://cedardb.com/blog/german_strings/

https://pola.rs/posts/polars-string-type/

Asides

What is a byte? Computers store information in 1s and 0s. Each 1 or 0 is a "bit." There are eight bits in a byte!

Four bits are in a nibble :)

Why did we choose a structure 16 bytes large to represent a string? It's just a magic upper bound; larger string structures can be slower. See https://cedardb.com/blog/strings_deep_dive/#function-call-optimizations.