Text mining
Tries

Marjahan Begum and Anders Kalhauge

cphbusiness

Spring 2017

Text
    Alphabets
    Strings and Languages
    Tries

An **alphabet** is

a finite set of characters

In our case it will also be

- ☐ ordered
- ☐ perfectly hashed

- English letters $\{A, B, \ldots, Z\}$
- Decimal digits $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- Binary digits $\{0, 1\}$
- Traffic light colours

What is meant by:

- ☐ Fininte set of characters
- ☐ Order of characters
- ☐ Perfect hashing of characters

cphbusiness
COPENHAGEN BUSINESS ACADEMY

In Java create functions, that can perfectly hash

☐ English letters, input is `char`

☐ digits, input is `int`

☐ red, yellow, green in that order, input is `String`

```java
public int indexOf(char letter) {
 ...
 }
```

A **string** is

a finite sequence of characters

from an alphabet
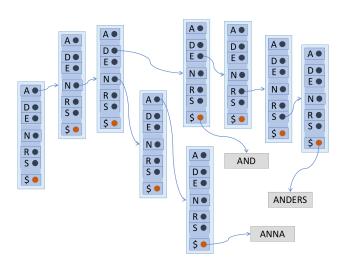
including the empty string ϵ

A **language** is

a set of strings

from an alphabet

including the empty set $\emptyset$