Breaking the Vigenère Cipher

Unknown Language

Duke
UNIVERSITY

Breaking Vigenére

- Step 3: Unknown Language
- For each language, need:
  - Word list
  - Most common character
- Can then try breaking
  - As before: most real words
  - Maximize word count across languages

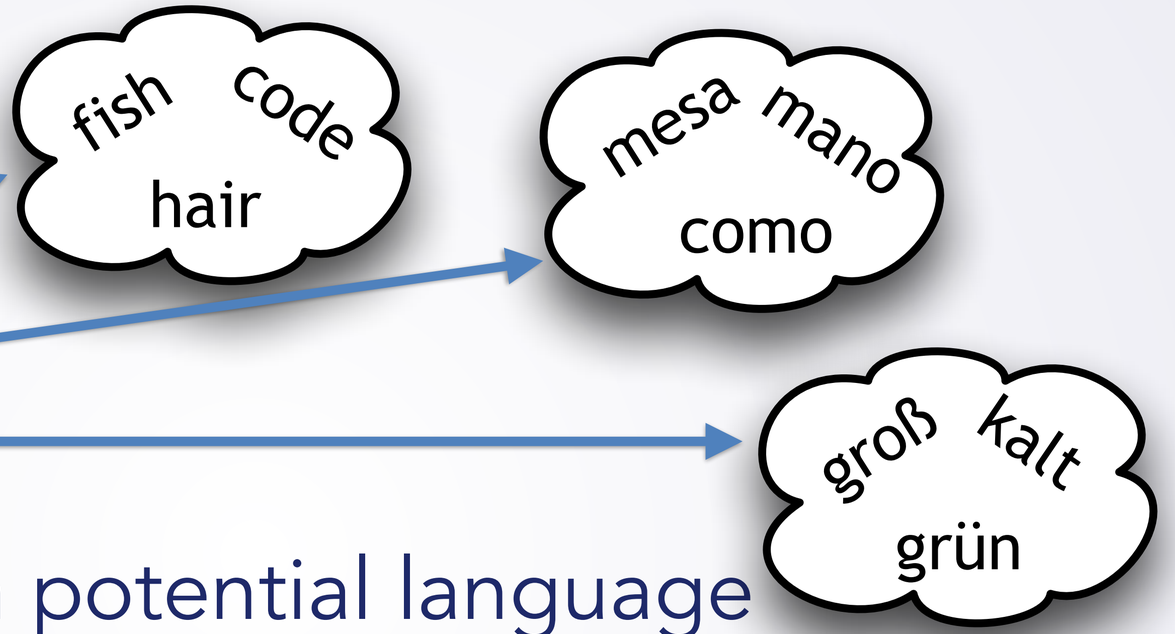
General Plan

Key	Value
English	
Spanish	
German	

- Read words for each potential language
  - Use `readDictionary` from before
- Make HashMap<String, HashSet<String>>
  - Key: language name
  - Value: result of `readDictionary`

General Plan

Key	Value
English	
Spanish	
German	



- Read words for each potential language
  - Use `readDictionary` from before
- Make HashMap<String, HashSet<String>>
  - Key: language name
  - Value: result of `readDictionary`

Composition: Complicated Types

HashMap<String, HashSet<String>>

- Remember: Composition
  - Can put pieces together

Plan for Coding

- Two new methods:
  - char mostCommonCharIn(HashSet<String> words)
  - public void breakForAllLangs(String encrypted, HashMap<String, HashSet<String>> languages)
- Modify two old methods:
  - public void breakVigenere ()
  - public String breakForLanguage(String encrypted, HashSet<String> dictionary)

Details of What to Do

```
char mostCommonCharIn(HashSet<String> words)
```

- ‘e’ not always most frequent
- Count frequency of letters in dictionary
- Proficient in counting occurrences, max

Details of What to Do

```
public void breakForAllLangs(String encrypted,
        HashMap<String, HashSet<String>> languages)
```

- Try each language in languages.keySet()
  - Use breakForLanguage
  - How many words did it end up with?
  - Pick best language + its decryption

Details of What to Do

```
public void breakVigenere ()
```

- Method you call from BlueJ
  - Read **all** languages' dictionaries
  - Call **breakForAllLangs** instead of **breakForLanguage**

Details of What to Do

public String breakForLanguage(String encrypted, HashSet<String> dictionary)

- Use mostCommonCharIn instead of ‘e’
  - Find right letter for this language