Web Tool for Phonemes Week-6

University at Buffalo The State University of New York

Agenda

- The Educator's Word Frequency guide
- Word Frequency - Kaggle (derived from Google Web Trillion Word Corpus)
- Vowels Articulation
- User Interface
- Kubernetes
- Fast-API
- Minimal and Maximal pair queues

The Educator's Word Frequency guide

- WFG has 60,527 samples of text obtained from 6333 textbooks, works of Literature and works of fiction and non-fiction used in schools and colleges.
- Researchers defined mathematical and statistical models to analyse the word frequencies.
- First Section: U>=1

  Statistical indices:

  SFI: Standard Frequency Index based on the total Corpus (log of U)

  D: Dispersion across content areas based on the total Corpus (words used in different subject areas)

  U: frequency expressed as N per million, weighted by D (frequency of the word per million)

  F: raw frequency based on the total Corpus
- Second Section: words with U < 1
- Third Section: words begin with ‘ or ` or numerals or abbreviations
- Fourth Section: unrounded U values for all words displayed in Section 1 and 2 in descending order

Word Frequency

This dataset contains the counts of the 333,333 most commonly-used single words on the English language web, as derived from the Google Web Trillion Word Corpus.

It uses n-gram frequency analysis.

In our dataset,

200k words out of 240k words have the word frequency.

Source : https://www.kaggle.com/datasets/rtatman/english-word-frequency/

https://drive.google.com/file/d/1mv8XXv2R7oBg4V7eZmhlkok_Fzr0tnzy/view?usp=share_link

the	23135851162
of	13151942776
and	12997637966
to	12136980858
a	9081174698
in	8469404971
for	5933321709
is	4705743816
on	3750423199
that	3400031103

Vowel Articulation Data

Height

Backness

Rounding


```
{
  "IPA_Number": 308,
  "Symbol": "u",
  "Symbol_Name": "lower-case U",
  "Description": "",
  "Voicing": "-",
  "Place_of_Articulation": "-",
  "Manner_of_Articulation": "-",
  "Height": "close",
  "Backness": "back",
  "Rounding": "rounded",
  "Status": "",
  "UCS_Code": "0075",
  "AFII_Code": "E265"
},
```

Vowel Articulation Data

```
{
    "phoneme": ["AE", "P", "AH", "L"],
    "POA":         ["-", "bilabial", "-", "dental;alveolar"],
    "MOA": ["-", "plosive", "-", "lateral approximant"],
    "VOA": ["-", "voiceless", "-", "voiced"],
    "H": ["near-open", "-", "open-mid", "-"],
    "B": ["front", "-", "back", "-"],
    "R": ["unrounded", "-", "unrounded", "-"]
}
```

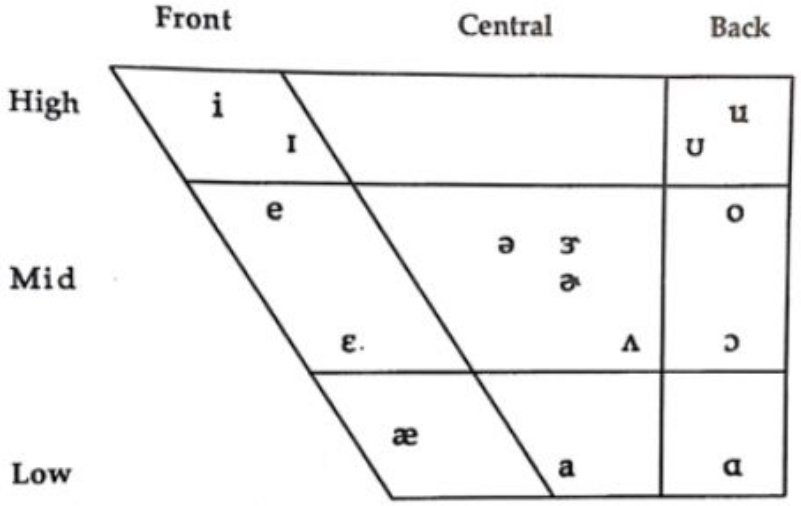


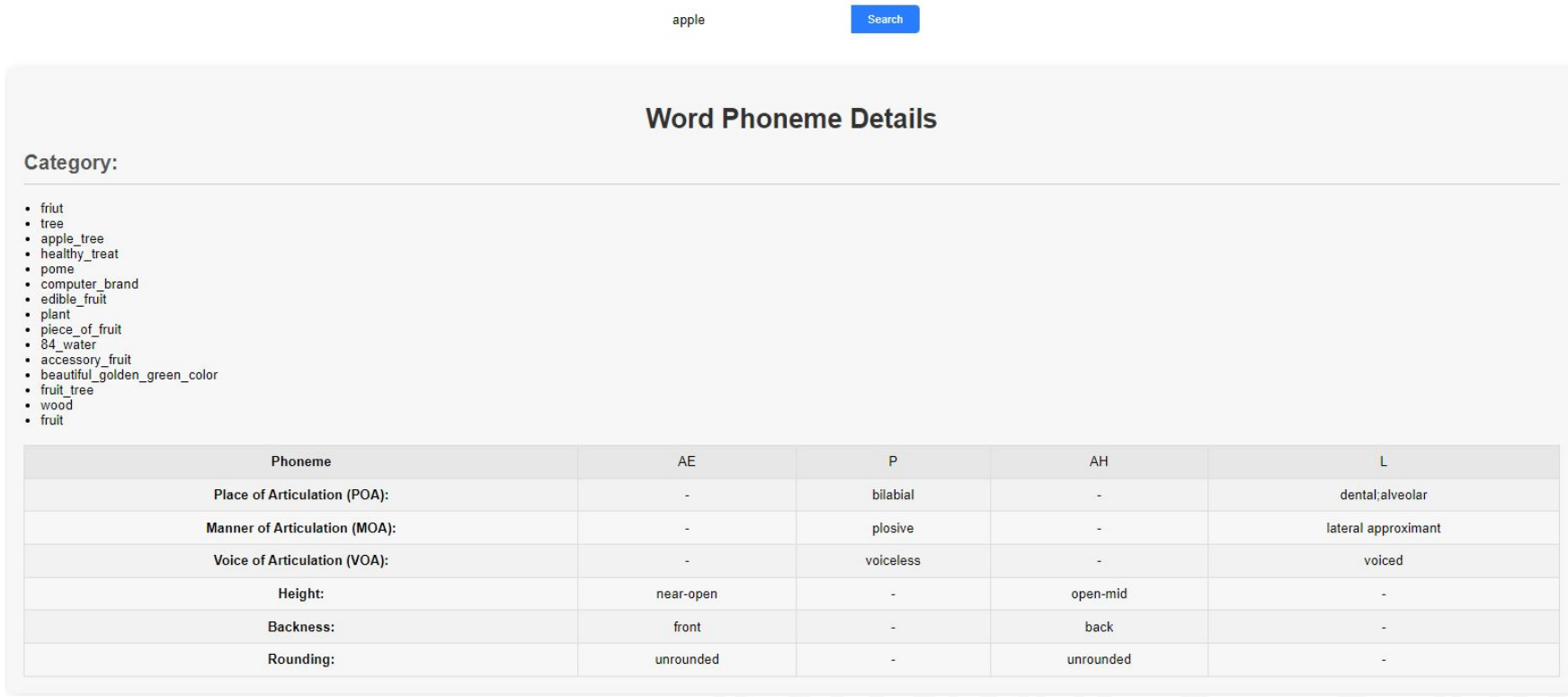
FIGURE 4.1 The vowel quadrilateral for American English vowels.

User Interface

- Used React to build UI.

Components:

- Word details
- SearchBar
- Index.js



Phoneme	AE	P	AH	L
Place of Articulation (POA):	-	bilabial	-	dental;alveolar
Manner of Articulation (MOA):	-	plosive	-	lateral approximant
Voice of Articulation (VOA):	-	voiceless	-	voiced
Height:	near-open	-	open-mid	-
Backness:	front	-	back	-
Rounding:	unrounded	-	unrounded	-

Kubernetes

USED pods for Solr, Nodejs and Fast API

and wrote respective service

FAST-API

Minimal and Maximal Pair Queries

- Minimal and Maximal Pairs if A Category and length of the phoneme is given
- Minimal and Maximal Pairs If a word is given
- Minimal and Maximal Pairs If a Category, and Two differing phonemes are given