

Date: 30-08-24

→ AI snake oil → fraud

- How to identify AI snake oil
eg. claims, exaggeration

Reading local files :-

```
f.open ('document.txt')
```

→ to write in the text.

```
f.open ('', 'w')  
f.write ('Hello')
```

→ write mode

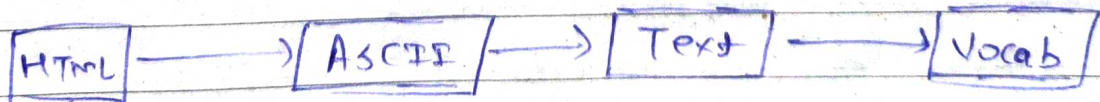
Similarly

'a' ⇒ for append

'r' ⇒ for read only

'w' ⇒ to write (overwrites previous content)

~~##~~ ## NLP Pipeline :-



Regular Expression :- tool to find patterns from a text

eg. words with 'ing'

eg.

wordlist = [w for w in nltk.corpus.words.words('en')
if w.islower()]

↓

all words in english

eg. word ending with 'ing' :-

regular
expression

symbol for
ending with

stuff = [w for w in wordlist if re.search('ing\$', w)]

for start with => ('^win', w)

→ string :- to find smallest unmodified part of word

• string to list → .split()

• list to string → .join()

Nltk chapter-4 :-

- assignment
- equality (comparisons)

• Questions of style :-

→ Procedural vs Declarative

• Variable scope

⇒ Algorithm design :-

- How to design to work with lesser resources but should be fast enough.

i.e.

efficient algorithms prioritized over inefficient.

eg.

Bubble sort vs Merge sort

Collocation & tagging words [chapter-5] :-

→ Part of speech tagging :- tagging words in the given sentence as part of speech.

- word itself & its location inside the sentence itself ~~gives~~ serves as features.

• word-tokenize \Rightarrow separates all words of given sentence.

\rightarrow Default tokenizer

\rightarrow Regular Expression tokenizer :- given default tokenizer for words with certain pattern

\rightarrow The lookup tokenizer

unigram \rightarrow using 1-word info

bigram \rightarrow using 2-word info