Date:- 30-08-24

→ AI snake oil → fraud

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

Reading local files :-

f. open ('document.text')

→ to write in the text. → 'write mode'

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

Similarly 'a' => for append
'r' => for read only
'w' => to write (overwrites previous content)

NLP Pipeline :-

HTML → ASCII → Text → Vocab

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' :-

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

regular expression

symbol for ending with

for start with ⇒ `('^win', w)`

→ <u>Stemming</u> :- to find smallest unmodified part of word

- string to list → `.split()`
- list to string → `.join()`

Nltk chapter-4 :-

- assignment
- equality (comparisions)

- 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 priortized over inefficient.

eg.

Bubble sort vs merge sort

Categorizing & tagging words [chapter-5] :-

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

- coord itself & its location inside the sentences itself ~~gives~~ serves as features.

  • word_tokenize => seperates all words of given sentence

→ Default tagger

→ Regular Expression tagger :- given default tags for words with certain pattern

→ The lookup tagger

  unigram → using 1-word info
  bigram → using 2-word info