→ Edit Distance :- (Spell correct" / String similarity)

kitten → sitting

. s i t t i n g

.

.

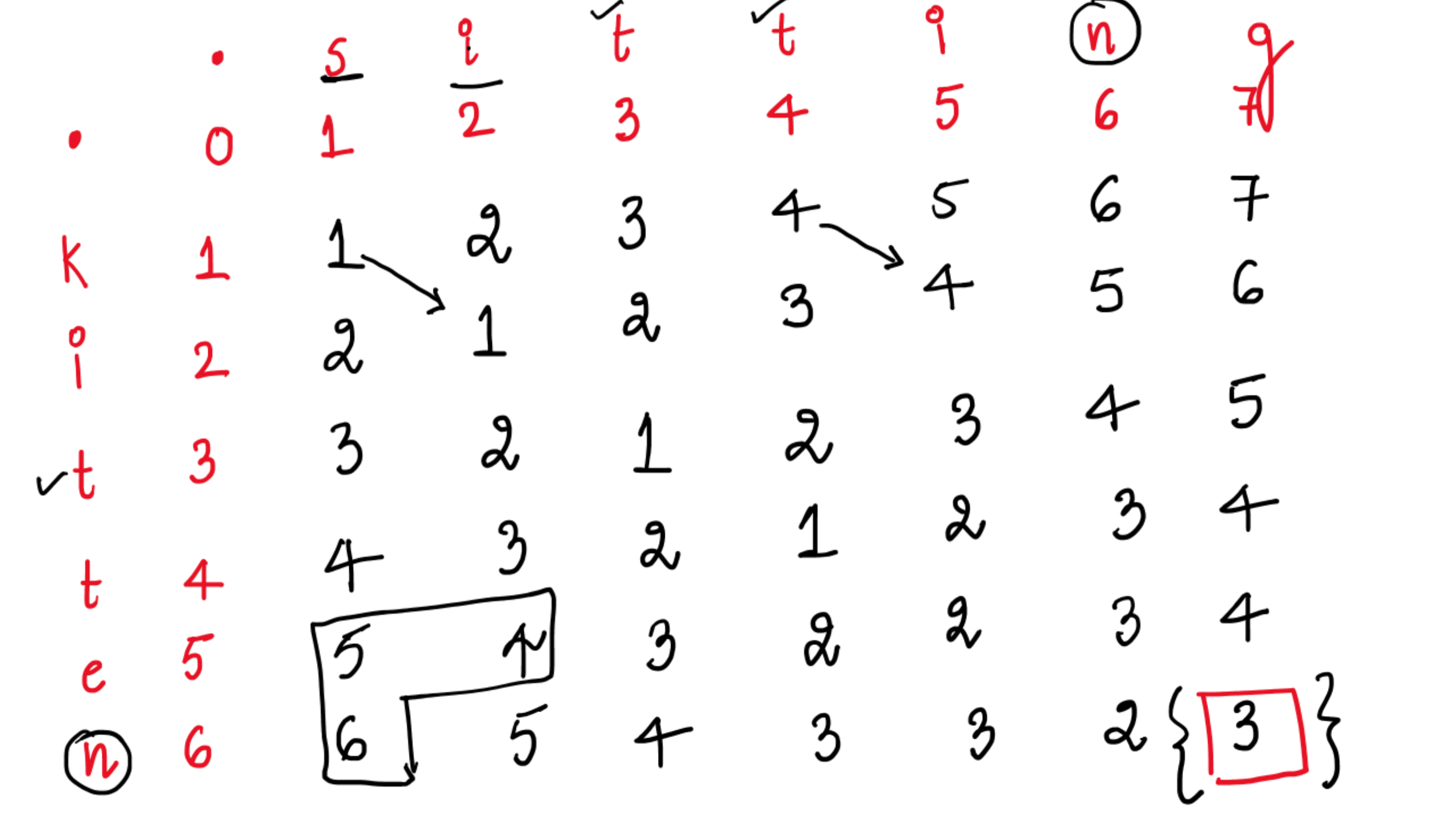
if r == c :

    ele = diagonal

else: (r != c)

    ele = min(all) + 1

    0 + 1 = 1


. 0 1 2 3 4 5 6
s 1 2 3 4 5 6 7
i 2 3 4 5 6 7
t 3 4 5 6 7
t 4 5 6 7
i 5 6 7
n 6 7
q 7
. 0 1 2 3 4 5 6
k 1 2 3 4 5 6
i 2 1 2 3 4 5
t 3 2 1 2 3 4
t 4 3 2 1 2 3
i 5 4 3 2 2 3
n 6 5 4 3 3 2
3

s k

i i

t 3 t

t t

i e

n n

g -

① (Substitut$^{n}$)

No ← Max { P(eating | sitting) ✓ , P(kitten | sitting) ✓ }

No

No

② (Substitution)

No

③ (Insertion)

3 operations

sitting → 3 → eating ✓

sitting → 3 → kitten

① closeness (Edit dist

② Word frequency

→ Probabilistic Approach:

C → correct word

w

* Bayes theorem: (Conditional Prob)

$$P(A/B) = P(C/w) = \frac{\{P(w/c)\} \times P(c)}{P(w)}$$

P(c) → frequency

$P_1 = P(\text{eating}/\text{sitting})$

$P_2 = P(\text{kitten}/\text{sitty})$

$P_n = P(\text{n}/\text{sitty})$

→ final word is with max prob.

→ Part of Speech (POS): { Syntactic → adjective

Sent = 'Upgrad is an edtech organisation'

Upgrad → Proper Noun
is → Helping verb
an → Determinant
edtech → adjective
organisation → Common Noun

} POS tags

* We do have pre-trained model for POS & NER

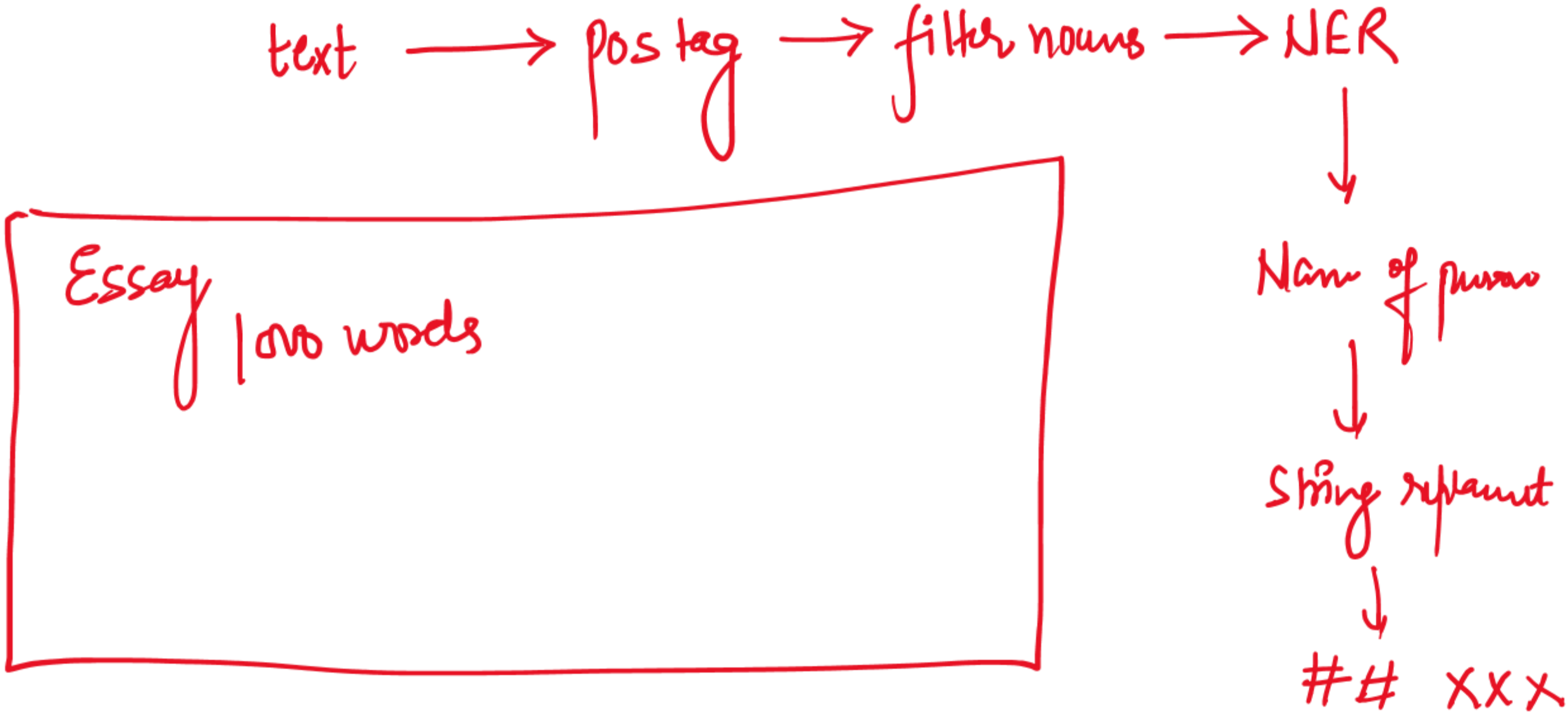
→ NER tags: Named entity Recognition

↓

further classificat$^{n}$ of noun

text ⟶ Sol$^{n}$ → text with mask names

- Name of person → Name of medicine etc etc
- Name of org
- Name of city


text → pos tag → filter nouns → NER
Essay
1000 words
Name of person
String replace
##XXX

Topic modelling

Text Summarisation
(initial version)

Dependency tags

Essay

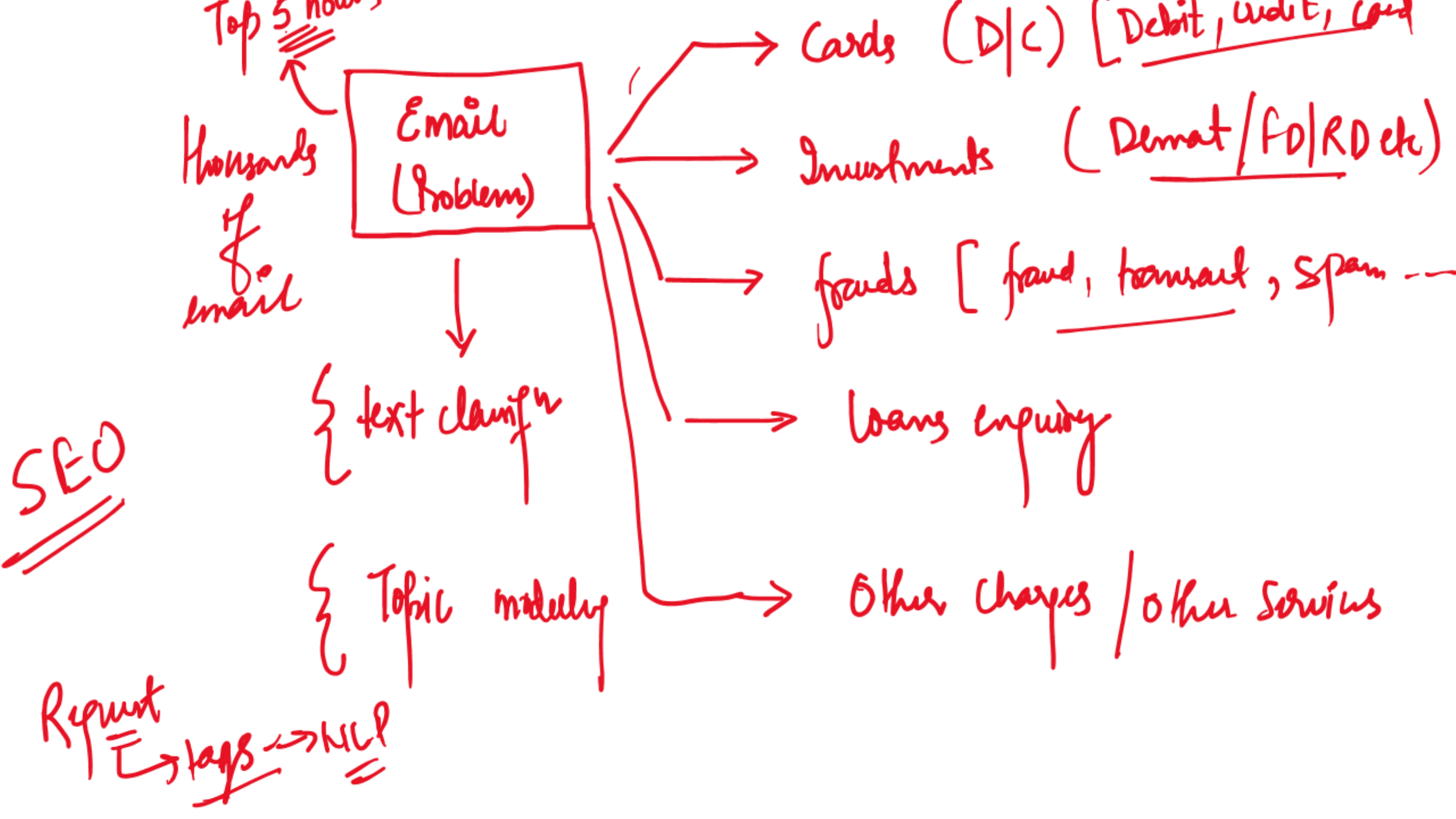
↓

POS tag

↓

filter Nouns → frequency distribution ✓

{ USA 10
politics 4
Barack Obama 2 }


Top 5 Issues
Thousands of email
Email (Problem)
Cards (D|C) [Debit, Credit, Card
Investments (Demat/FD|RD etc)
frauds [fraud, transact, Spam
loans enquiry
Other charges / other Services
text classfn
Topic modeling
SEO
Report
tags → NLP