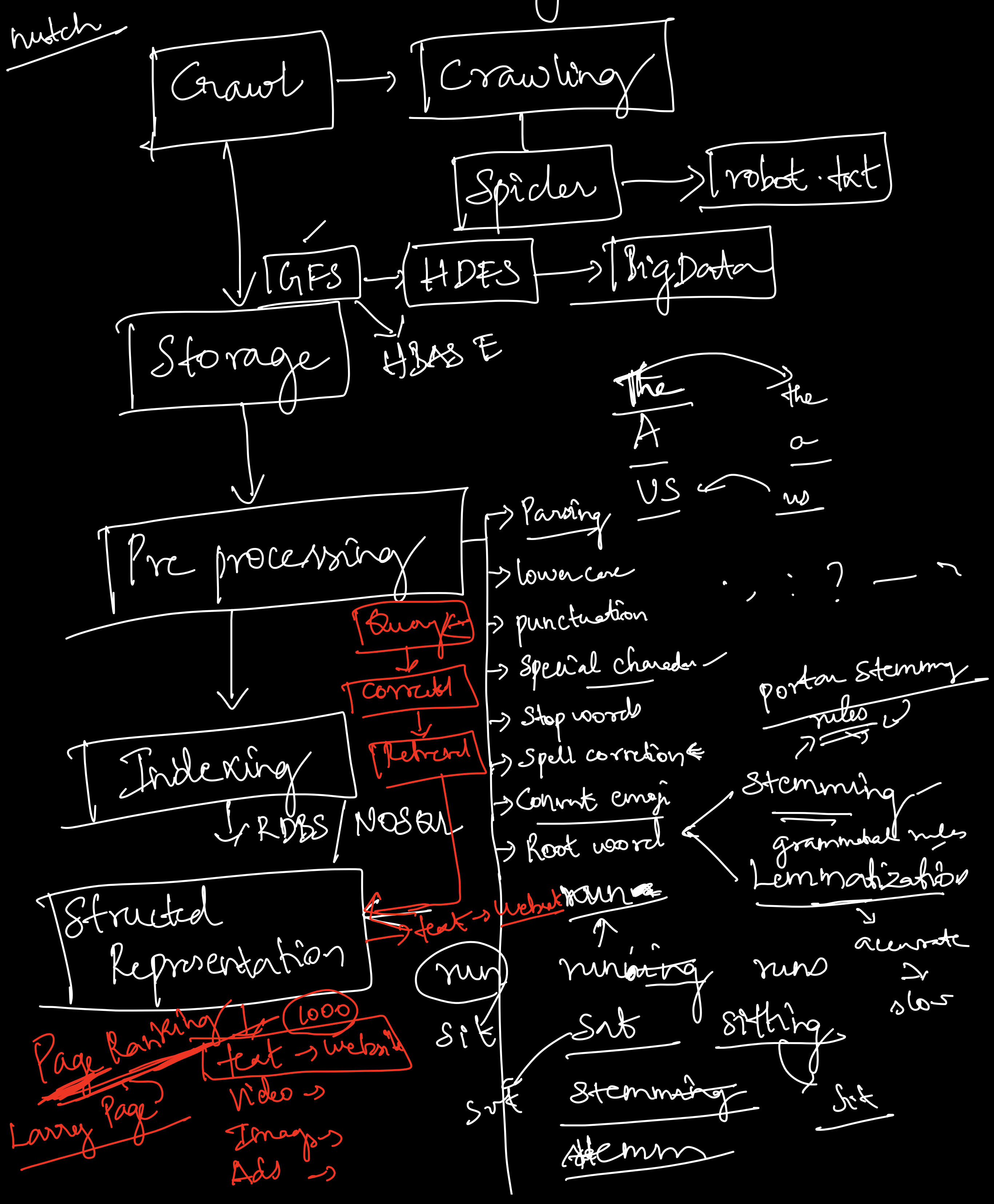
Search Engine



→ Tokenization

I had enough breafest ⇒ enough breakfast

breakfast was awesome ⇒ breakfast awesome

I had to take a trip ⇒ take trip

I went for a running ⇒ go run

he runs very fast → run very fast

he runs fast, while he ran he collapsed — run fast run collabs

n grams ⇒ Uni gram, bi gram, tri gram

[illegible]

	Uni gram	bi gram
enough breakfast →	[enough, breakfast)	[enough breakfas)
breakfast awesome →	[breakfast, awesomi]	[breafs aw)
take trip →	[take, trip]	[take trip)
go run →	[go, run]	(go rn)
run very fast →	[run, fast)	[run very] [very fast

	enough	breakfast	awesome	take	trip	run	enough breakf	breakaw
1.	1	1	0	0	0		1	0
2	0	1	1	0	0		0	1
3.								
4.								
5.								
6								

Indexing

1. <u>One hot</u> encoding →
   a. Occurances <u>not</u> there for documents and corpus
   b. orders are not preserved
   c. Context is not presered

2. Bag of Words
   a. Occurrences not there for corpus
   b. orders are not preserved
   c. Context is not ther

3. TF-IDF

4. Word embedding