Surga kiron NLP - ASSIGNMENT -1 25/09/23 consider or resurption الا Mutate good turing discovery. en: Your she findly and now want the principle was the scrood turing Interition: . We the worst of things which are seen one to help estimate the consist of temps never seen. · Stimilarly, we count of toings which occur 'et times to estimate the count of teings which occur 'e' times · let we be the no. of things that occur is times. in prequent of it. 70位: 18(14. ) 101=3 : 102=1 ; 113=1 ; 11-11:11 · MEE [mowmum likelihood attention] went for Ne is 'e' where As, good turing extimete hintion of N car in the til is will not (? (23m) [8] (mes) c = (20) New The was the tree that of in a state ner en Re-attentes the probability of low court agrains by considering agreems we lighter weinte, the provider a discounted probability walker. · Algo, estimates the probability of zero went things by considering aprome by want 1. Dote: Ne is the court of tuings we've prente times. (2 x = (2 x 1) ALCA 1 8111=8 83 W Self Median a family so IN R IVI

22201030

us consider on example:

en: You are fieling and have caregor to following files:

- · corp: 10
  - · perch: 3
  - · Awlitefish: 2
  - trout : 1
  - for (mon : 1
  - cel: 1

Total: 18 (14. ; N = 3 ; N2 = 1; N3 = 1; N0 = 10;

26 in the authority out for Mc a 'c' when it, How likely is It touch next proces is tolong? I stant a popular know

1/18 (wee) " = (cri) Now

How they is it that next species is a new specie sullies confish?

(3 poerely, we say 1 0/18 p. but ung our inherror to estimate trings we now

once to extimete the new things, we get: wi/N=3/18 (: N, = 3)

so cultivates the presenting of som out things by considering agreed it is it is Then, the likelihood of the next species being e solmon will be discounted runny the below warrens

More the is the want of things me've then L3 | c\* = (c+i) NC+1 MCG P=1/18

ex (salma) = (+1) NH =

 $\frac{2\times1}{3}$ 1. Pur (Calmin) = 2/2/18

: Mary Hall

2	. What are the inner we tay indeterminery and tokenization?
	The inner we tag indeterminey U tokerremons can be summerred a follows:
M	-> Tag Indeterminay:
3-00,000	-> Tag indeterminay:
· Surpen	· Ambigusty between Madiple toys:
	Last probable from our represent
	Tay indeterminery occurs when a work we be executed w/ multiple por tags,
	end it is deallerging to disembiguest which tag is correct.
	The second of the model to the second
5. 7	Herdling Indeterminate tags:
	There is a series to the serie
2 3	some tagger allow the use of multiple tags for indeterminate words,
main of	In some copie tay on placed we apple has on one tay is chosen
	duing training.
	· Common indeterminate tage:
	Enamples of woman tag indetermine ares include distinguishing to be
at lon	and duiding whether a word a exadicative or nown as prenomice
	and deciding whether a word a set adjustive or nown as presente
THE MILLS	
	The said of the said of
Tra V	Day - may be soon who we have
	The state of the s
	and moved to health the sum when

## · Tokenization:

· Role in Word Splitting: Los pulsuration pur la made all so toril. . S.

Tokenization is enected for splitting words and determining words boundaries.

Josevaple, in the Renn Treckelle at Aritish National Confus, constrouting and possessive fame on superested from their terms dury to receivables.

you suppressed and min

Word splitting Differences:

Different expose may tokenne words differently for mutance, the

pen Treebook trust multiple uports words (eig bow love) es e

reparete words, while other tagnets with CI Hagnet, in trust

then we imple words.

and the Cokhan Woals:

Most be tagging algorithmy of proper dicherates to only Pol treps
to words. However, the emotion of proper name, accompying and
new woman nound or verbs occur regularly, and the toossers
needs metabols to hardle the inknown words.