➢ **Example:**

- Compute the probability of sentences:
  - I want English food
  - I want Chinese food

- Bigram counts table for eight of the words (out of V = 1446) in the Berkeley Restaurant Project corpus of 9332 sentences.

	I	want	to	eat	Chinese	food	lunch	Spend
I	5	827	0	9	0	0	0	2
want	2	0	608	1	6	6	5	1
To	2	0	4	686	2	0	6	211
Eat	0	0	2	0	16	2	42	0
Chinese	1	0	0	0	0	82	1	0
Food	15	0	15	0	1	4	0	0
Lunch	2	0	0	0	0	1	0	0
Spend	1	0	1	0	0	0	0	0

- Unigrams Counts:

I	want	to	Eat	Chinese	food	lunch	Spend
2533	927	2417	746	158	1093	341	278

- Bigram probabilities after normalization (dividing each row by the unigram counts above):

	I	want	To	eat	Chinese	food	lunch	Spend
I	0.002	0.33	0	0.0036	0	0	0	0.00079
want	0.0022	0	0.6	0.0011	0.0065	0.0065	0.0054	0.0011
To	0.00083	0	0.0017	0.28	0.00083	0	0.0025	0.087
Eat	0	0	0.0027	0	0.021	0.0027	0.056	0
Chinese	0.0063	0	0	0	0	0.052	0.0063	0
Food	0.014	0	0.014	0	0.00092	0.0037	0	0
lunch	0.0059	0	0	0	0	0.0029	0	0
spend	0.0036	0	0.0036	0	0	0	0	0

- we have a few other useful probabilities: Domain knowledge:

$P(I \mid \text{<s>}) = 0.25$ $P(\text{English} \mid \text{want}) = 0.0011$

$P(\text{food} \mid \text{English}) = 0.5$ $P(\text{</s>} \mid \text{food}) = 0.68$

$$P(\text{<s> I want English food </s>}) = P(I|\text{<s>}) * P(\text{want} \mid I) *$$
$$P(\text{English} \mid \text{want}) * P(\text{food} \mid \text{English}) * P(\text{</s>}|\text{food})$$
$$= 0.25 * 0.33 * 0.0011 * 0.5 * 0.68$$
$$= 0.000031$$

$$P(\text{<s> I want Chinese food </s>}) = P(I|\text{<s>}) \times P(\text{want} \mid I) *$$
$$P(\text{Chinese} \mid \text{want}) * P(\text{food} \mid \text{Chinese}) * P(\text{</s>}|\text{food})$$
$$= 0.25 * 0.33 * 0.0065 * 0.052 * 0.68$$
$$= 0.000019$$

- **Example:**

<start> I	0.25	Want some	0.04
<start> I'd	0.06	Want Thai	0.01
<start> Tell	0.04	To eat	0.26
<start> I'm	0.02	To have	0.14
I want	0.32	To spend	0.09
I would	0.29	To be	0.02
I don't	0.08	British food	0.60
I have	0.04	British restaurant	0.15
Want to	0.65	British cuisine	0.01
Want a	0.05	British lunch	0.01
Eat British	0.001	Eat Chinese	0.02
Chinese food	0.60	Food <end>	0.01

I want to eat British food

P(I want to eat British food)

= P(<start>I want to eat British food<end>)

= P(I|<start>) * P(want | I) * P(to | want) * P(eat | to) *

P(British | eat) * P(food | British) * P(<end> | food)

= 0.25 * 0.32 * 0.65 * 0.26 * 0.001 * 0.6 * 0.01

= 0.0000008112

I want to eat Chinese food

P(I want to eat Chinese food)

= P(<start>I want to eat Chinese food<end>)

= P(I|<start>) * P(want | I) * P(to | want) * P(eat | to) *

P(Chinese | eat) * P(food | Chinese) * P(<end> | food)

= 0.25 * 0.32 * 0.65 * 0.26 * 0.02 * 0.6 * 0.01

= 0.0000016

I want some food

Assume: P(food | some) = 0.02

P(I want some food)

= P(<start>I want some food<end>)

= P(I|<start>) * P(want | I) * P(some | want) *

P(food | some) * P(<end> | food)

= 0.25 * 0.32 * 0.04 *0.02* 0.01

= 0.00000064

➢ Example:

- classify docs into spam or not spam

	system's prediction	correct answer	TP	FP	FN	TN
d1	→ Y	N		1		
d2	→ Y	Y	1			
d3	→ N	Y			1	
d4	→ N	N				1
d5	→ Y	N		1		

$$Precision = \frac{TP}{TP+FP} = \frac{1}{1+2} = \frac{1}{3} = 0.333$$

$$Recall = \frac{TP}{TP+FN} = \frac{1}{1+1} = \frac{1}{2} = 0.5$$

$$F = \frac{2 * 0.333 * 0.5}{0.333 + 0.5} = 0.4$$

$$accuracy = \frac{TP+TN}{All}$$

$$= \frac{1+1}{5} = 0.4$$

➢ Example:

given

Doc 1 : Information Retrieval Systems
Doc 2 : Information Storage
Doc 3 : Digital Speech Synthesis Systems
Doc 4 : Speech Filtering, Speech Retrieval

$$idf = \log \frac{N}{df_t}$$

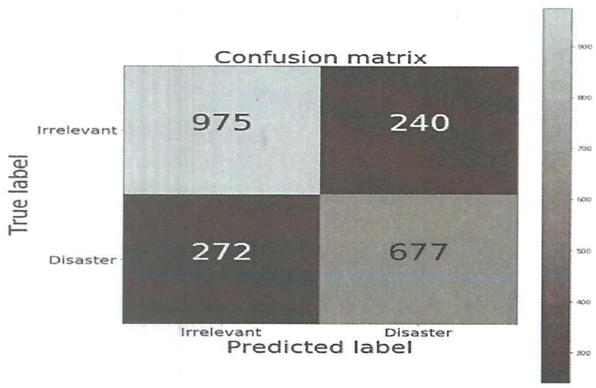
Term	IDF	TF			
		Doc 1	Doc 2	Doc 3	Doc 4
Digital	log 4	0	0	$\frac{1}{4}$	0
Filtering	log 4	0	0	0	$\frac{1}{4}$
Information	log 2	$\frac{1}{3}$	$\frac{1}{2}$	0	0
Retrieval	log 2	$\frac{1}{3}$	0	0	$\frac{1}{4}$
Speech	log 2	0	0	$\frac{1}{4}$	$\frac{1}{2}$
Storage	log 4	0	$\frac{1}{2}$	0	0
Synthesis	log 4	0	0	$\frac{1}{4}$	0
Systems	log 2	$\frac{1}{3}$	0	$\frac{1}{4}$	0

TF-idf

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Confusion Matrix				
		Predicted		Total
		Yes T	No F	
Actual	Yes T	6954 TP	46 FN	7000
	No F	412 FP	2588 TN	3000
Total		7366	2634	10000



1. Ngram can be used in
   a. Spelling error
   b. Word detection
   c. POS tagging
   d. All of the above

2. The application that detect names of people or locations in text
   a. Name detection
   b. Name recognition
   c. Named Entity recognition
   d. Concept Entity recognition

3. … in NLP is many ways to say the same
   a. **Redundancy**
   b. Ambiguity
   c. Overfitting
   d. Underfitting

4. … in NLP is many senses of the same data
   a. Redundancy
   b. **Ambiguity**
   c. Overfitting
   d. Underfitting

5. … means the study of patterns of speech sounds.
   a. **Phonology**
   b. Morphology
   c. Syntax
   d. semantics

6. … means how words can be changed by inflection or derivation
   a. Phonology
   b. **Morphology**
   c. Syntax
   d. Semantics

7. the order of the words and structure between them are the role of ...
   a. Phonology
   b. Morphology
   c. **Syntax**
   d. Semantics

8. ... means the study of how meaning is created by words and phrases.
   a. Phonology
   b. Morphology
   c. Syntax
   d. **Semantics**

9. How to recognize speech and how to synthesize it is the interest of ....
   e. **Phonology**
   f. Morphology
   g. Syntax
   h. Semantics

10.... means the study of meaning in contexts.
   a. **Pragmatics**
   b. Discourse
   c. Syntax
   d. Semantics

11.... the coherent (logical) groups of sentences
   a. Pragmatics
   b. **Discourse**
   c. Syntax
   d. Semantics

12."every man loves a woman" has .... ambiguity
   a. **semantic**
   b. syntax
   c. phonological
   d. lexical

13. "I shot an elephant in my pyjamas" has .... ambiguity

   a. semantic
   b. **syntax**
   c. phonological
   d. lexical

14. the word "dance" has .... ambiguity

   a. semantic
   b. syntax
   c. phonological
   d. **lexical**

15. "I made her duck" has .... ambiguity

   a. semantic
   b. syntax
   c. lexical
   d. **all of the above**

16. ..... finding the boundaries of sentences in text.

   a. Sentence segmentation
   b. tokenization
   c. POS tagging
   d. Semantic analysis

17. "New York" is a problem in ...

   a. Sentence segmentation
   b. tokenization
   c. POS tagging
   d. Semantic analysis

18. Handling abbreviations and proper nouns is a necessary step in ...

   a. Sentence segmentation
   b. tokenization
   c. POS tagging
   d. Semantic analysis

19.“Mr., Dr.” is a problem in …

a. **Sentence segmentation**
b. tokenization
c. POS tagging
d. Semantic analysis

20.Can, May , Must are … and are considered in … class

a. Modal Verbs, open
b. **Modal Verbs, closed**
c. Closed, Modal Verbs
d. None of the above

21.Ngram approach is used as … POS tagging

a. Rule based
b. **Stochastic**
c. Transformation based
d. Decision tree

22.HMM approach is used as … POS tagging

a. Rule based
b. **Stochastic**
c. Transformation based
d. Decision tree

23. the grammatical form of a noun or pronoun

a. **Case**
b. Person
c. Number
d. Gender

24.All the following is considered closed classes except

a. **get**
b. on
c. the
d. up

25.POS tagging is useful for all the following tasks except

a. **Tokenization**
b. Text to speech
c. Lemmatization
d. Parsing

26.All the following is needed for Rule bases tagging except

a. Dictionary
b. List of potential POS
c. Rules
d. **Large Corpora**

27.The main issue of word frequency approach in POS tagging is ...

a. Dictionary
b. **Inadmissible sequence of tags**
c. Rules
d. Large Corpora

28."They" is considered ... person

a. First
b. Second
c. **Third**
d. No

29."We" is considered ... person

a. **First**
b. Second
c. Third
d. No

30."You" is considered ... person

a. First
b. **Second**
c. Third
d. No

31. In … The tag encountered most frequently with the word in the training set is the one assigned to an ambiguous instance of that word

   a. **Word frequency approach**
   b. Tag sequence approach
   c. Rule based approach
   d. Katz

32. … performs better on small training sets

   a. Katz
   b. **Simple interpolation (Jelinek-Mercer)**
   c. Absolute discounting
   d. Kneser-Ney

33. … perform better on large training sets

   a. **Katz**
   b. Simple interpolation (Jelinek-Mercer)
   c. Absolute discounting
   d. Kneser-Ney

34. … performs well on N-grams with large counts

   a. **Katz**
   b. Simple interpolation (Jelinek-Mercer)
   c. Absolute discounting
   d. Kneser-Ney

35. … best on N-grams with small counts

   a. Katz
   b. Simple interpolation (Jelinek-Mercer)
   c. Absolute discounting
   d. **Kneser-Ney**

36. … models are superior on … models for low (nonzero) counts

   a. **Interpolation, bakoff**
   b. Backoff, interpolation
   c. Word frequency
   d. None of the above

37.... can deal with many words in a dictionary

a. unigram
b. bigram
c. lookup
d. word frequency

38.... how many bigrams in the sentence "I want English food"

a. 5
b. 6
c. 7
d. 8

39.... represents that many ways of saying the same

a. Redundancy
b. Ambiguity
c. Structured
d. Unstructured

40.... represents that many senses of the same data

a. Redundancy
b. Ambiguity
c. Structured
d. Unstructured

41.... data is information with an organized structure such as relational databases

a. Redundancy
b. Ambiguity
c. Structured
d. Unstructured

42. is not in regular databases and growing exponentially making up the most of real world data

a. Redundancy
b. Ambiguity
c. Structured
d. Unstructured

43. Spell checking, Keyword search, finding synonymous are examples of … NLP applications
    a. **Easy**
    b. Medium
    c. Hard
    d. Other

44. Extracting info from websites, classifying Text, Sentiment Analysis are examples of **…** NLP applications
    a. Easy
    b. **Medium**
    c. Hard
    d. Other

45. Machine translation, spoken dialogue systems (chatbots), Complex Question-answering, Semantic Analysis, Co-reference/ Anaphora resolution are examples of **…** NLP applications
    a. Easy
    b. Medium
    c. **Hard**
    d. Other