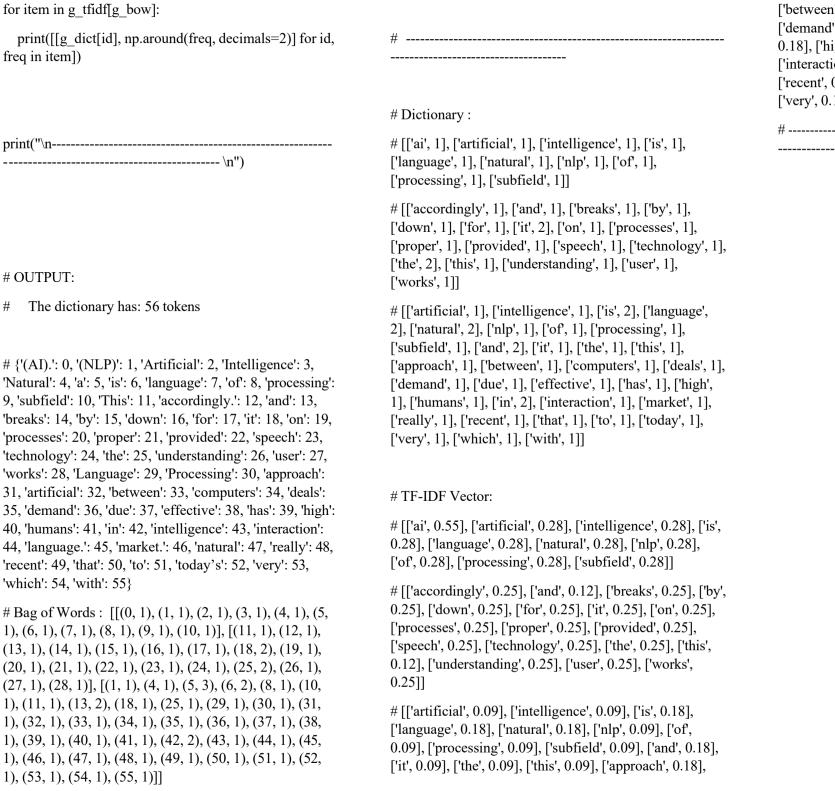
#######################################	# [[('are', 1), ('hello', 1), ('how', 1), ('you', 1)], [('how', 1), ('you', 1), ('doi, 2)], [('are', 2), ('you', 3), ('doing', 2), ('hey', 1), ('what', 2), ('yes', 1)]]	"Natural Language Processing (NLP) is a subfield of artificial intelligence that deals with the interaction between computers and humans in natural language."
import gensim		
import pprint	#Assignment no : 2	tokens1 = [[item for item in line.split()] for line in
from gensim import corpora	#Name: Ishwar wagh	doc_list]
from gensim.utils import simple_preprocess	#Batch : B4	g_dict1 = corpora.Dictionary(tokens1)
doc_list = [
"Hello, how are you?", "How do you do?", "Hey what are you doing? yes you What are you doing?"	#Roll no : 6565	<pre>print("The dictionary has: " +str(len(g_dict1)) + " tokens\n")</pre>
	#Title: Natural Language Processing (NLP) using Gensim	
	#TFID	print(g_dict1.token2id)
]	,	
<pre>doc_tokenized = [simple_preprocess(doc) for doc in doc_list]</pre>	import gensim	g_bow =[g_dict1.doc2bow(token, allow_update = True) for token in tokens1]
dictionary = corpora.Dictionary()	import pprint	print("Bag of Words: ", g_bow)
BoW_corpus = [dictionary.doc2bow(doc, allow_update=True) for doc in doc_tokenized]	from gensim import corpora, models	
	from gensim.utils import simple_preprocess	print("\n
print(BoW_corpus)	import numpy as np	\n")
id_words = [[(dictionary[id], count) for id, count in line] for line in BoW_corpus]	from nltk.tokenize import sent_tokenize, word_tokenize	<pre>g_dict = corpora.Dictionary([simple_preprocess(line) for line in doc_list])</pre>
print(id_words)	import gensim from gensim.models import Word2Vec	g_bow = [g_dict.doc2bow(simple_preprocess(line)) for line in doc_list]
# OUTPUT		maint("Distingues")
# [[(0, 1), (1, 1), (2, 1), (3, 1)], [(2, 1), (3, 1), (4, 2)], [(0, 2), (3, 3), (5, 2), (6, 1), (7, 2), (8, 1)]]	doc_list = [print("Dictionary : ")
	"Natural language processing (NLP) is a subfield of	for item in g_bow:
	Artificial Intelligence (AI).",	<pre>print([[g_dict[id], freq] for id, freq in item])</pre>
	"This technology works on the speech provided by the user breaks it down for proper understanding and processes it accordingly.",	g_tfidf = models.TfidfModel(g_bow, smartirs='ntc')
	"This is a very recent and effective approach due to which it has a really high demand in today's market."	<pre>print("\nTF-IDF Vector:")</pre>



['between', 0.18], ['computers', 0.18], ['deals', 0.18], ['demand', 0.18], ['due', 0.18], ['effective', 0.18], ['has', 0.18], ['high', 0.18], ['humans', 0.18], ['in', 0.36], ['interaction', 0.18], ['market', 0.18], ['really', 0.18], ['recent', 0.18], ['that', 0.18], ['to', 0.18], ['today', 0.18], ['very', 0.18], ['which', 0.18], ['with', 0.18]]