In [1]:
```
from nltk.tokenize import word_tokenize
```

In [2]:
```
file=open('opinion.txt','r')
text=file.read()
```

In [3]:
```
w=word_tokenize(text)
len(w)
```

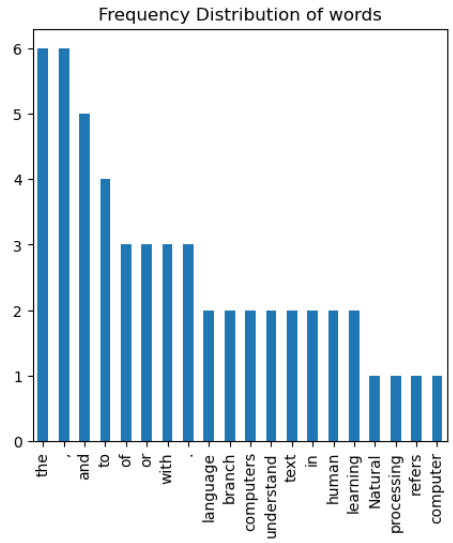
Out[3]:
```
93
```

In [4]:
```
from nltk.probability import FreqDist
afd=FreqDist(w).most_common(20)
print(afd)
```

```
[('the', 6), (',', 6), ('and', 5), ('to', 4), ('of', 3), ('or', 3), ('with', 3), ('.', 3), ('language', 2), ('branch', 2), ('computers', 2), ('understand', 2), ('text', 2), ('in', 2), ('human', 2), ('learning', 2), ('Natural', 1), ('processing', 1), ('refers', 1), ('computer', 1)]
```

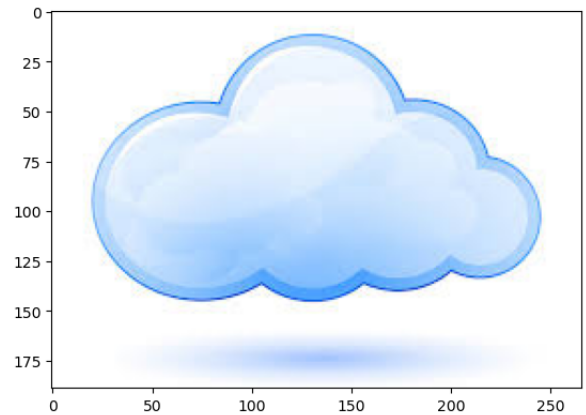
In [13]:
```
import matplotlib.pyplot as plt
import pandas as pd
afd=pd.Series(dict(afd))
fig,ax=plt.subplots(figsize=(5,5))
afd.plot(kind='bar')
plt.title('Frequency Distribution of words')
```

Out[13]:
```
Text(0.5, 1.0, 'Frequency Distribution of words')
```

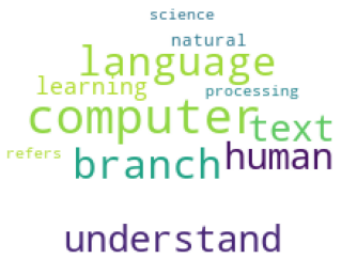


In [20]:
```
from skimage.io import imread
tree=imread('cloud.jpg')
plt.imshow(tree)
```

Out[20]:
```
<matplotlib.image.AxesImage at 0x2bb91173190>
```



In [22]:
```
from wordcloud import WordCloud,STOPWORDS
import matplotlib.pyplot as plt
s=set(STOPWORDS)
w=WordCloud(width=800,height=800,stopwords=s,background_color='white',min_font_size=10,mask=tree).generate(text)
plt.figure(figsize=(5,5),facecolor=None)
plt.imshow(w)
plt.axis('off')
plt.show()
```



In [8]:
```
import nltk
from nltk.metrics.distance import edit_distance
```

In [9]:
```
nltk.download('words')
from nltk.corpus import words
cw=words.words()
```

```
[nltk_data] Error loading words: <urlopen error [WinError 10061] No
[nltk_data]     connection could be made because the target machine
[nltk_data]     actively refused it>
```

In [10]:
```
iw=['happpy','amzzzzzzzzzzzzzzing','inteliegent','cllege']
for word in iw:
    temp=[(edit_distance(word,w),w)for w in cw if w[0]==word[0]]
    print(sorted(temp,key=lambda val:val[0])[0][1])
```

```
happy
amazing
intelligent
college
```

In [11]:
```
text=text.upper()
print(text)
```

```
NATURAL LANGUAGE PROCESSING REFERS TO THE BRANCH OF COMPUTER SCIENCE AND MORE SPECIFICALLY, THE BRANCH OF ARTIFICIAL INTELLIGENCE OR AI, CONCERNED WITH GIVING COMPUTERS THE ABI
LITY TO UNDERSTAND TEXT AND SPOKEN WORDS IN MUCH THE SAME WAY HUMAN BEINGS CAN. NLP COMBINES COMPUTATIONAL LINGUISTICS WITH STATISTICAL, MACHINE LEARNING, AND DEEP LEARNING MOD
ELS. TOGETHER, THESE TECHNOLOGIES ENABLE COMPUTERS TO PROCESS HUMAN LANGUAGE IN THE FORM OF TEXT OR VOICE DATA AND TO UNDERSTAND ITS FULL MEANING, COMPLETE WITH THE SPEAKER OR
WRITERS INTENT AND SENTIMENT.
```

In [12]:
```
text=text.lower()
print(text)
```

```
natural language processing refers to the branch of computer science and more specifically, the branch of artificial intelligence or ai, concerned with giving computers the abi
lity to understand text and spoken words in much the same way human beings can. nlp combines computational linguistics with statistical, machine learning, and deep learning mod
els. together, these technologies enable computers to process human language in the form of text or voice data and to understand its full meaning, complete with the speaker or
writers intent and sentiment.
```

In [ ]: