```
# XML scrapping for xml sheet

import os
os.chdir(r"C:\Users\Hanshu\Downloads\14. NLP WEB SCRAPING\xml_single articles")

import xml.etree.ElementTree as ET

tree = ET.parse("769952.xml")
root = tree.getroot()

root = ET.tostring(root, encoding='utf8').decode('utf8')

root

import re, string, unicodedata
import nltk

from bs4 import BeautifulSoup
from nltk import word_tokenize, sent_tokenize
from nltk.corpus import stopwords
from nltk.stem import  LancasterStemmer, WordNetLemmatizer
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt

def strip_html(text):
    soup = BeautifulSoup(text, "html.parser")
    return soup.get_text()

def remove_between_square_brackets(text):
    return re.sub(r'\[[^]]*\]', '', text)


def clean_text(text):
    text = strip_html(text)
    text = remove_between_square_brackets(text)
    text = re.sub(' ', "", text)
    return text

sample = clean_text(root)
print(sample)

wordcloud = WordCloud(width=800, height=400,
                      background_color='white',
                      stopwords=STOPWORDS,
                      min_font_size=10).generate(sample)


plt.figure(figsize=(10,5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```