

Alexander Guerra

Average length of webpages in tokens:----- 407.078125

Top ten most frequent email addresses found in the entire webpage collection:-----

[('ksugrad@kennesaw.edu', 25), ('ksuinfo@kennesaw.edu', 7), ('wellstaradvising@kennesaw.edu', 6), ('select@1.2', 5), ('hmattord@kennesaw.edu', 5), ('srodri51@kennesaw.edu', 5), ('bcoe_advising@kennesaw.edu', 5), ('cyberinstitute@kennesaw.edu', 4), ('ggavin@kennesaw.edu', 4), ('rjeansig@kennesaw.edu', 4)]

Percentage of webpages that contain at least one email address:----- 33.41346153846153 %

Top 30 most common words:----- [('and', 11618), ('the', 10102), ('of', 8652), ('to', 7231), ('in', 4902), ('a', 4428), ('for', 3475), ('Kennesaw', 2986), ('Campus', 2703), ('Resources', 2518), ('is', 2262), ('For', 2178), ('KSU', 2117), ('The', 2026), ('with', 1937), ('State', 1907), ('Marietta', 1886), ('GA', 1735), ('you', 1648), ('on', 1641), ('470-KSU-INFO', 1612), ('(470-578-4636)', 1609), ('that', 1515), ('your', 1438), ('be', 1426), ('by', 1409), ('will', 1392), ('are', 1386), ('students', 1350), ('or', 1287)]

Top 30 most common words with stop words removed:----- [('Kennesaw', 2986), ('Campus', 2703), ('Resources', 2518), ('KSU', 2117), ('State', 1907), ('Marietta', 1886), ('GA', 1735), ('470-KSU-INFO', 1612), ('(470-578-4636)', 1609), ('students', 1350), ('2023', 1055), ('College', 1021), ('University.', 909), ('Kennesaw.', 907), ('30144', 882), ('Contact', 858), ('South', 847), ('Media', 842), ('Marietta.', 841), ('@', 830), ('30060', 829), ('Info', 822), ('Chastain', 822), ('Road', 822), ('1100', 822), ('Links', 822), ('Rights', 821), ('1000', 817), ('Pkwy', 816), ('Phone', 813)]

```
Average length of webpages in tokens:----- 407.078125
Top ten most frequent email addresses found in the entire webpage collection:----- [('ksugrad@kennesaw.edu', 25), ('ksuinfo@kennesaw.edu', 7), ('wellstaradvising@kennesaw.edu', 6), ('select@1.2', 5), ('hmattord@kennesaw.edu', 5), ('srodri51@kennesaw.edu', 5), ('bcoe_advising@kennesaw.edu', 5), ('cyberinstitute@kennesaw.edu', 4), ('ggavin@kennesaw.edu', 4), ('rjeansig@kennesaw.edu', 4)]
Percentage of webpages that contain at least one email address:----- 33.41346153846153 %
Top 30 most common words:----- [('and', 11618), ('the', 10102), ('of', 8652), ('to', 7231), ('in', 4902), ('a', 4428), ('for', 3475), ('Kennesaw', 2986), ('Campus', 2703), ('Resources', 2518), ('is', 2262), ('For', 2178), ('KSU', 2117), ('The', 2026), ('with', 1937), ('State', 1907), ('Marietta', 1886), ('GA', 1735), ('you', 1648), ('on', 1641), ('470-KSU-INFO', 1612), ('(470-578-4636)', 1609), ('that', 1515), ('your', 1438), ('be', 1426), ('by', 1409), ('will', 1392), ('are', 1386), ('students', 1350), ('or', 1287)]
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\alex\AppData\Roaming\nltk data...
[nltk_data] Package stopwords is already up-to-date!
Top 30 most common words with stop words removed:----- [('Kennesaw', 2986), ('Campus', 2703), ('Resources', 2518), ('KSU', 2117), ('State', 1907), ('Marietta', 1886), ('GA', 1735), ('470-KSU-INFO', 1612), ('(470-578-4636)', 1609), ('students', 1350), ('2023', 1055), ('College', 1021), ('University.', 909), ('Kennesaw.', 907), ('30144', 882), ('Contact', 858), ('South', 847), ('Media', 842), ('Marietta.', 841), ('@', 830), ('30060', 829), ('Info', 822), ('Chastain', 822), ('Road', 822), ('1100', 822), ('Links', 822), ('Rights', 821), ('1000', 817), ('Pkwy', 816), ('Phone', 813)]
```

