DR. ALVIN'S PUBLICATIONS

HOW TO PERFORM SUPER SIMPLE TEXT ANALYSIS

DR. ALVIN ANG

<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th></tr>
<tr><td>1</td><td>RANK</td><td>FREQ</td><td>WORD</td><td>% OF TOTAL</td><td>VIZ</td></tr>
<tr><td>2</td><td>1</td><td>102</td><td>THE</td><td>6.59767141</td><td>||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>3</td><td>2</td><td>57</td><td>TO</td><td>3.686934023</td><td>|||||||||||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>4</td><td>3</td><td>49</td><td>THAT</td><td>3.169469599</td><td>|||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>5</td><td>4</td><td>36</td><td>IN</td><td>2.328589909</td><td>||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>6</td><td>5</td><td>30</td><td>OF</td><td>1.940491591</td><td>||||||||||||||||||||||||||||||</td></tr>
<tr><td>7</td><td>6</td><td>29</td><td>AND</td><td>1.875808538</td><td>|||||||||||||||||||||||||||||</td></tr>
<tr><td>8</td><td>7</td><td>29</td><td>IS</td><td>1.875808538</td><td>|||||||||||||||||||||||||||||</td></tr>
<tr><td>9</td><td>8</td><td>24</td><td>I</td><td>1.552393273</td><td>||||||||||||||||||||||||</td></tr>
<tr><td>10</td><td>9</td><td>23</td><td>A</td><td>1.48771022</td><td>|||||||||||||||||||||||</td></tr>
<tr><td>11</td><td>10</td><td>22</td><td>IT</td><td>1.423027167</td><td>||||||||||||||||||||||</td></tr>
<tr><td>12</td><td>11</td><td>20</td><td>CHINA</td><td>1.293661061</td><td>||||||||||||||||||||</td></tr>
<tr><td>13</td><td>12</td><td>19</td><td>HAVE</td><td>1.228978008</td><td>|||||||||||||||||||</td></tr>
<tr><td>14</td><td>13</td><td>19</td><td>NUMBER</td><td>1.228978008</td><td>|||||||||||||||||||</td></tr>
<tr><td>15</td><td>14</td><td>17</td><td>NOT</td><td>1.099611902</td><td>|||||||||||||||||</td></tr>
<tr><td>16</td><td>15</td><td>16</td><td>WE</td><td>1.034928849</td><td>||||||||||||||||</td></tr>
<tr><td>17</td><td>16</td><td>15</td><td>WILL</td><td>0.9702457956</td><td>|||||||||||||||</td></tr>
<tr><td>18</td><td>17</td><td>14</td><td>WOULD</td><td>0.9055627426</td><td>||||||||||||||</td></tr>
<tr><td>19</td><td>18</td><td>13</td><td>THIS</td><td>0.8408796895</td><td>|||||||||||||</td></tr>
<tr><td>20</td><td>20</td><td>12</td><td>THEY</td><td>0.7761966365</td><td>||||||||||||</td></tr>
</table>

CONTENTS

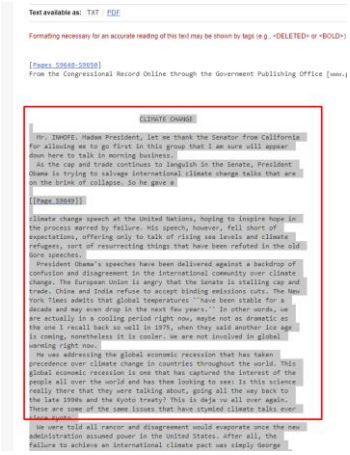
INTRODUCTION

- This article follows Bajak (2015).
- https://www.storybench.org/how-to-do-super-simple-textual-analysis/
- It makes use of a Web Frequency Indexer found here:
- https://www.lextutor.ca/freq/eng/
- We may expect an Excel output like this:

	A	B	C	D	E
1	RANK	FREQ	WORD	% OF TOTAL	VIZ
2	1	102	THE	6.59767141	
3	2	57	TO	3.686934023	
4	3	49	THAT	3.169469599	
5	4	36	IN	2.328589909	
6	5	30	OF	1.940491591	
7	6	29	AND	1.875808538	
8	7	29	IS	1.875808538	
9	8	24	I	1.552393273	
10	9	23	A	1.48771022	
11	10	22	IT	1.423027167	
12	11	20	CHINA	1.293661061	
13	12	19	HAVE	1.228978008	
14	13	19	NUMBER	1.228978008	
15	14	17	NOT	1.099611902	
16	15	16	WE	1.034928849	
17	16	15	WILL	0.9702457956	
18	17	14	WOULD	0.9055627426	
19	18	13	THIS	0.8408796895	
20	20	12	THEY	0.7761966365	

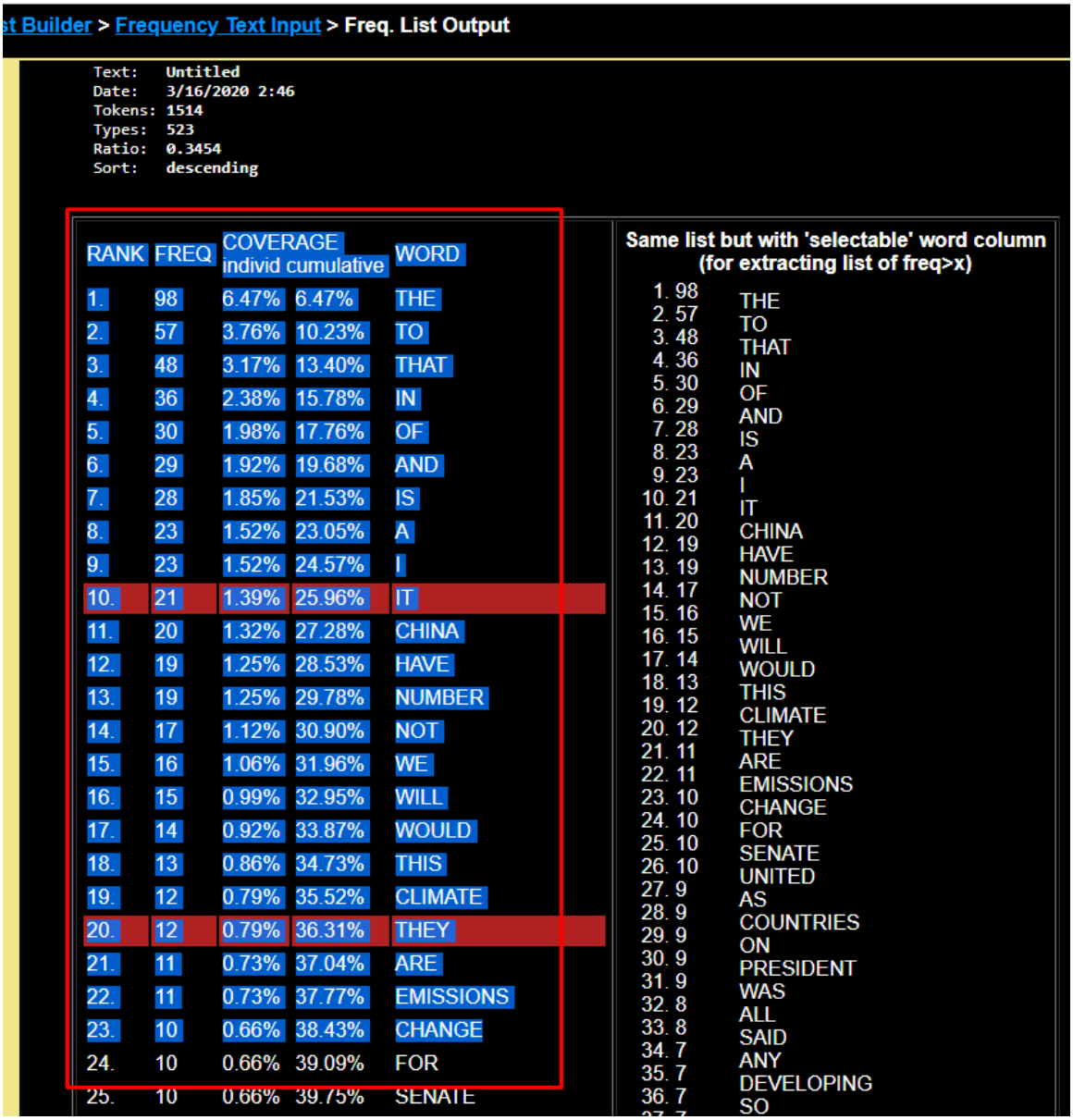
STEP 1: COPY PASTE TEXT

- Go to https://www.lextutor.ca/freq/eng/



- Go to https://www.congress.gov/congressional-record/2009/09/22/senate-section/article/s9648-2/
- Select the text and copy.
- Paste it into the 'Web Frequency Indexer' box → Click 'Submit Window'.

STEP 2: WEB FREQUENCY INDEXER OUTPUT



t Builder > Frequency Text Input > Freq. List Output

Text: Untitled
Date: 3/16/2020 2:46
Tokens: 1514
Types: 523
Ratio: 0.3454
Sort: descending

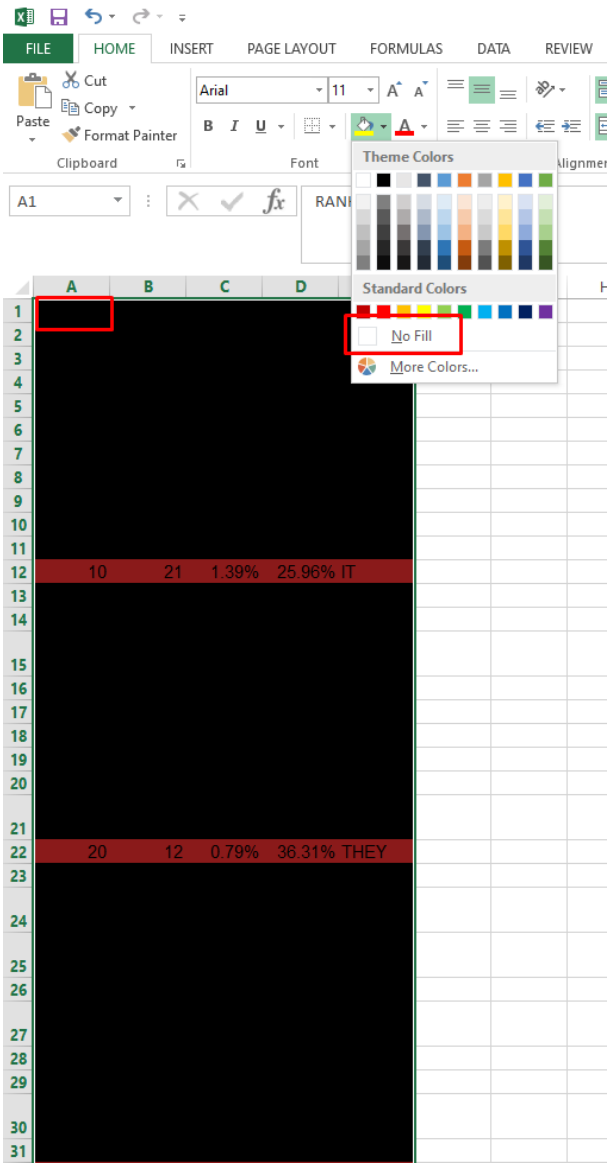
RANK	FREQ	COVERAGE individ	COVERAGE cumulative	WORD
1.	98	6.47%	6.47%	THE
2.	57	3.76%	10.23%	TO
3.	48	3.17%	13.40%	THAT
4.	36	2.38%	15.78%	IN
5.	30	1.98%	17.76%	OF
6.	29	1.92%	19.68%	AND
7.	28	1.85%	21.53%	IS
8.	23	1.52%	23.05%	A
9.	23	1.52%	24.57%	I
10.	21	1.39%	25.96%	IT
11.	20	1.32%	27.28%	CHINA
12.	19	1.25%	28.53%	HAVE
13.	19	1.25%	29.78%	NUMBER
14.	17	1.12%	30.90%	NOT
15.	16	1.06%	31.96%	WE
16.	15	0.99%	32.95%	WILL
17.	14	0.92%	33.87%	WOULD
18.	13	0.86%	34.73%	THIS
19.	12	0.79%	35.52%	CLIMATE
20.	12	0.79%	36.31%	THEY
21.	11	0.73%	37.04%	ARE
22.	11	0.73%	37.77%	EMISSIONS
23.	10	0.66%	38.43%	CHANGE
24.	10	0.66%	39.09%	FOR
25.	10	0.66%	39.75%	SENATE

Same list but with 'selectable' word column (for extracting list of freq>x)

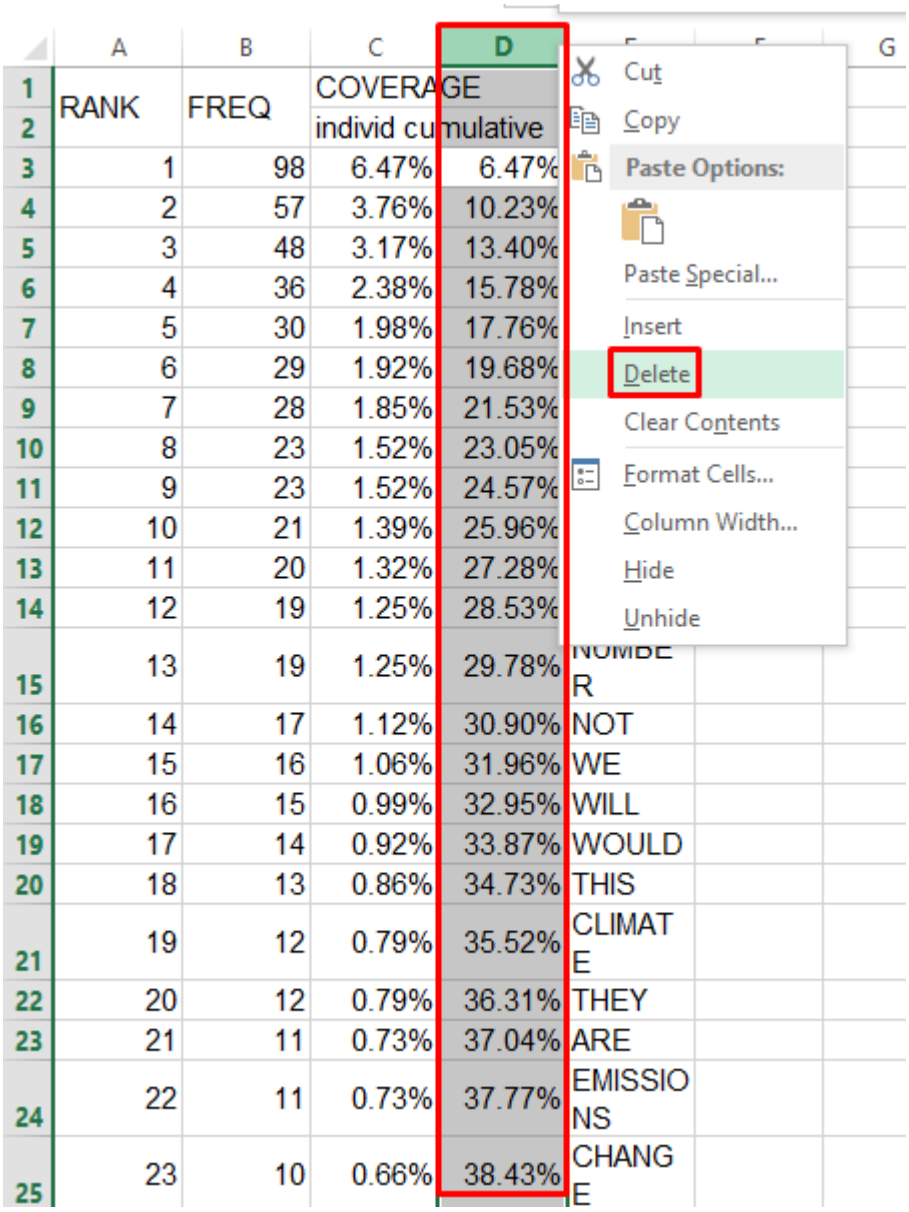
1. 98 THE
2. 57 TO
3. 48 THAT
4. 36 IN
5. 30 OF
6. 29 AND
7. 28 IS
8. 23 A
9. 23 I
10. 21 IT
11. 20 CHINA
12. 19 HAVE
13. 19 NUMBER
14. 17 NOT
15. 16 WE
16. 15 WILL
17. 14 WOULD
18. 13 THIS
19. 12 CLIMATE
20. 12 THEY
21. 11 ARE
22. 11 EMISSIONS
23. 10 CHANGE
24. 10 FOR
25. 10 SENATE
26. 10 UNITED
27. 9 AS
28. 9 COUNTRIES
29. 9 ON
30. 9 PRESIDENT
31. 9 WAS
32. 8 ALL
33. 8 SAID
34. 7 ANY
35. 7 DEVELOPING
36. 7 SO

- You will come to this new page.
- Select and copy the LHS of the table i.e. Rank / Freq / Coverage / Word.

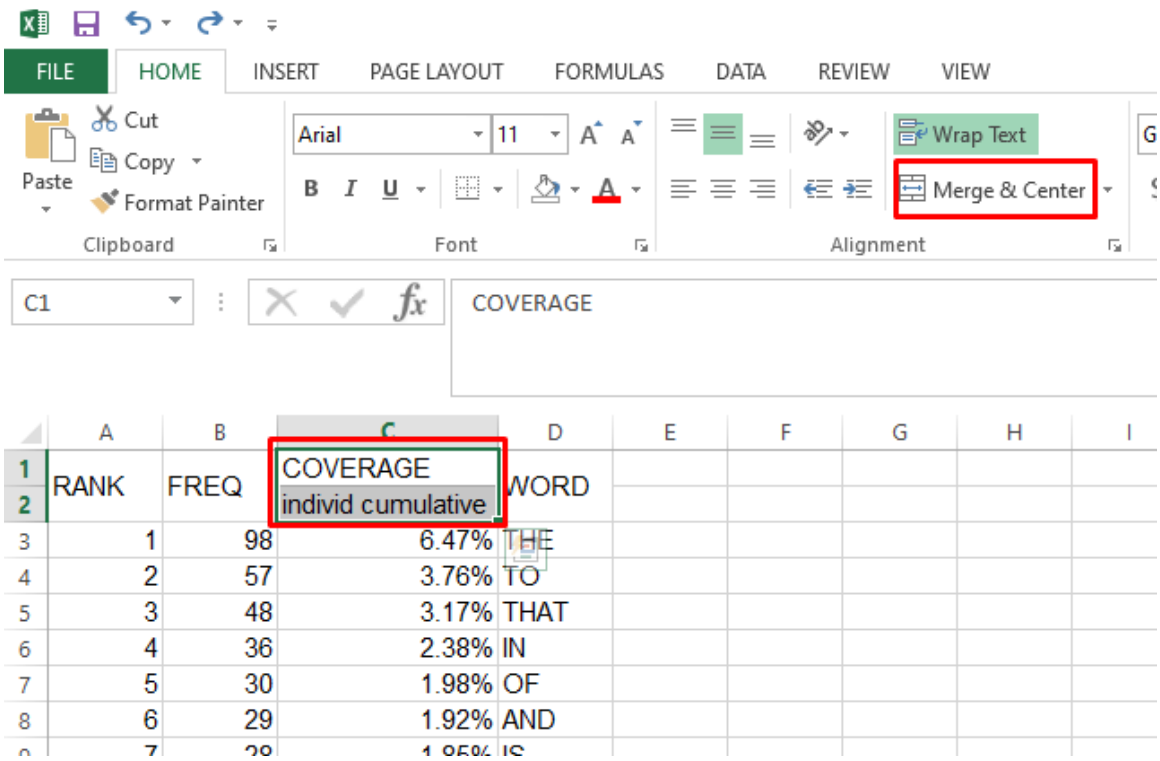
STEP 1: PASTING THE DATA



- Open Excel
- Paste everything into cell A1.
- Remember to change the background color to → No Fill.

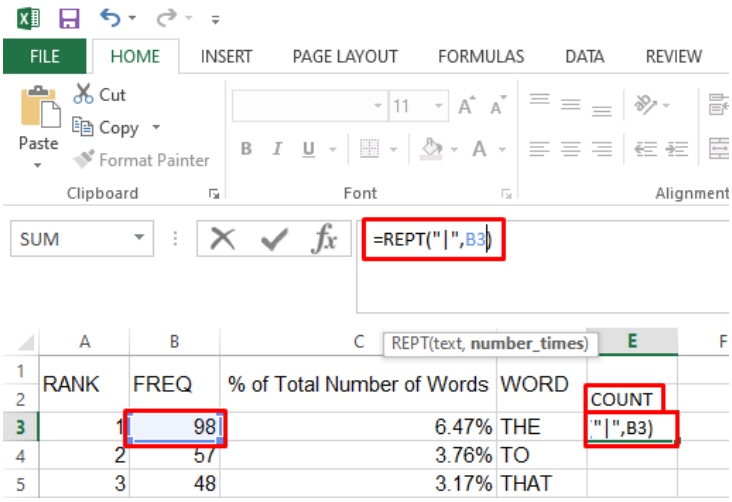


- Delete column D (for simplicity).

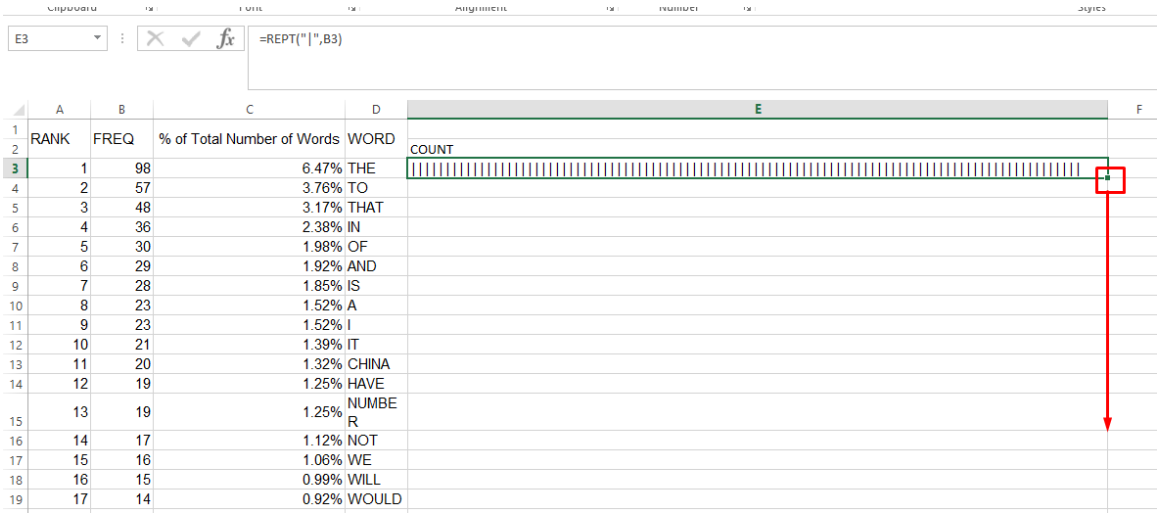


- Merge and Center the cell 'Coverage' and 'individ cumulative'.
- Change the title to "% of Total Number of Words".

STEP 2: CREATING THE COUNT " | "



- Create a new column called "Count"
- In the first row, type in
    - **= REPT ("|", B3)**
- Press Enter.



- It will draw many ||||||
- Drag it all the way down.

CONCLUSION

<table>
<tr><th>RANK</th><th>FREQ</th><th>% of Total Number of Words</th><th>WORD</th><th>COUNT</th></tr>
<tr><td>1</td><td>98</td><td>6.47%</td><td>THE</td><td>||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>2</td><td>57</td><td>3.76%</td><td>TO</td><td>|||||||||||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>3</td><td>48</td><td>3.17%</td><td>THAT</td><td>||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>4</td><td>36</td><td>2.38%</td><td>IN</td><td>||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>5</td><td>30</td><td>1.98%</td><td>OF</td><td>||||||||||||||||||||||||||||||</td></tr>
<tr><td>6</td><td>29</td><td>1.92%</td><td>AND</td><td>|||||||||||||||||||||||||||||</td></tr>
<tr><td>7</td><td>28</td><td>1.85%</td><td>IS</td><td>||||||||||||||||||||||||||||</td></tr>
<tr><td>8</td><td>23</td><td>1.52%</td><td>A</td><td>|||||||||||||||||||||||</td></tr>
<tr><td>9</td><td>23</td><td>1.52%</td><td>I</td><td>|||||||||||||||||||||||</td></tr>
<tr><td>10</td><td>21</td><td>1.39%</td><td>IT</td><td>|||||||||||||||||||||</td></tr>
<tr><td>11</td><td>20</td><td>1.32%</td><td>CHINA</td><td>||||||||||||||||||||</td></tr>
<tr><td>12</td><td>19</td><td>1.25%</td><td>HAVE</td><td>|||||||||||||||||||</td></tr>
<tr><td>13</td><td>19</td><td>1.25%</td><td>NUMBER</td><td>|||||||||||||||||||</td></tr>
<tr><td>14</td><td>17</td><td>1.12%</td><td>NOT</td><td>|||||||||||||||||</td></tr>
<tr><td>15</td><td>16</td><td>1.06%</td><td>WE</td><td>||||||||||||||||</td></tr>
<tr><td>16</td><td>15</td><td>0.99%</td><td>WILL</td><td>|||||||||||||||</td></tr>
<tr><td>17</td><td>14</td><td>0.92%</td><td>WOULD</td><td>||||||||||||||</td></tr>
<tr><td>18</td><td>13</td><td>0.86%</td><td>THIS</td><td>|||||||||||||</td></tr>
<tr><td>19</td><td>12</td><td>0.79%</td><td>CLIMATE</td><td>||||||||||||</td></tr>
<tr><td>20</td><td>12</td><td>0.79%</td><td>THEY</td><td>||||||||||||</td></tr>
<tr><td>21</td><td>11</td><td>0.73%</td><td>ARE</td><td>|||||||||||</td></tr>
<tr><td>22</td><td>11</td><td>0.73%</td><td>EMISSIONS</td><td>|||||||||||</td></tr>
</table>

- You will finish with what you see above.

REFERENCES

Bajak, A. (2015). "How to do super simple textual analysis." from https://www.storybench.org/how-to-do-super-simple-textual-analysis/.

ABOUT THE AUTHORS

Dr. Alvin Ang earned his Ph.D., Masters and Bachelor degrees from NTU, Singapore. He is a scientist, entrepreneur, as well as a personal/business advisor. More about him at www.AlvinAng.sg.