STOP LIST

1. the,189775
2. of,104955
3. and,87638
4. in,74237
5. to,66944
6. a,57369
7. is,39564
8. for,27776
9. as,26188
10. by,21430
11. that,20466
12. are,19637
13. with,18393
14. on,16107
15. be,15666
16. from,15663
17. or,15433
18. was,13155
19. it,12014
20. s,11778
21. an,11457
22. at,10955
23. which,10172
24. this,10044
25. has,9706
26. can,9284
27. have,8839
28. also,8408
29. used,7280
30. such,6989
31. more,6765
32. its,6601
33. not,6473
34. may,4929
35. one,4898
36. but,4635
37. these,4551
38. some,4548
39. they,4147
40. first,4139
41. all,3946
42. there,3942
43. will,3746
44. when,3689
45. many,3686
46. over,3512
47. us,3470
48. would,3148
49. through,3140
50. only,3024

Above is the minimal list of stop words filtered from unigram file. I have identified 50 stop words within top 83 words from the unigram file (sorted in decreasing order of term frequency). These words consist of common determiners, coordinating conjunctions, prepositions. I have chosen 50 because the words succeeding that are relevant and cannot be ignored.

filtered words

energy,17382  
power,11376  
heat,5888  
gas,5116  
fuel,4934  
solar,4833  
system,4779

Above words appear in top 83 from the unigram file, and these words have been excluded from stop list. These words could uplift the topical relevance for sustainable energy related searches although they have high term frequency. These could be domain specific depending upon the user query.