Brian Howe

CS 141, Lab Thurs. 4pm

Analyze the content of inaugural speeches and display data.

I was able to finish most of the assignment. I believe the code is for the most part correct but do to the differences in which the speeches were formated the code failed to work on some. This is obvious in the paragraphs column where some speeches have 1 when they clearly should have many. One place where I believe the code is to blame and not the format is the singular pronouns. For some reason all the counts for this are off and I don't know why.

Output: gets a little messy when copied to document but still readable. Running the program will print a much cleaner version.

Speaker | Date | Letrs | Words | Sents |

----------------------------------------------------------------------------

1. Abraham Lincoln | March 4 1861 | 16947 | 3628 | 135 |

2. | March 4 1865 | 3127 | 695 | 26 |

----------------------------------------------------------------------------

3. Andrew Jackson | March 4 1829 | 5549 | 1112 | 25 |

4. | March 4 1833 | 5745 | 1162 | 30 |

----------------------------------------------------------------------------

5. Barack Obama | January 20 2009 | 10631 | 2391 | 122 |

6. | January 21 2013 | 9434 | 2090 | 91 |

----------------------------------------------------------------------------

7. Benjamin Harrison | March 4 1889 | 21296 | 4350 | 157 |

----------------------------------------------------------------------------

8. Calvin Coolidge | March 4 1925 | 19402 | 4032 | 196 |

----------------------------------------------------------------------------

9. Dwight Eisenhower | January 20 1953 | 11000 | 2445 | 119 |

10. | January 21 1957 | 7074 | 1636 | 96 |

----------------------------------------------------------------------------

11. Franklin Pierce | March 4 1853 | 16363 | 3315 | 104 |

----------------------------------------------------------------------------

12. Franklin Roosevelt | January 20 1937 | 8419 | 1766 | 95 |

13. | January 20 1941 | 5896 | 1303 | 74 |

14. | January 20 1945 | 2318 | 529 | 29 |

15. | March 4 1933 | 8731 | 1856 | 86 |

----------------------------------------------------------------------------

16. George Bush | January 20 1989 | 9688 | 2320 | 153 |

----------------------------------------------------------------------------

17. George W Bush | January 20 2001 | 7046 | 1591 | 96 |

18. | January 20 2005 | 9473 | 2069 | 100 |

----------------------------------------------------------------------------

19. George Washington | April 30 1789 | 7044 | 1422 | 23 |

20. | March 4 1793 | 631 | 133 | 4 |

----------------------------------------------------------------------------

21. Grover Cleveland | March 4 1885 | 8246 | 1659 | 44 |

----------------------------------------------------------------------------

22. Grover cleveland | March 4 1893 | 10087 | 1987 | 58 |

----------------------------------------------------------------------------

23. Harry Truman | January 20 1949 | 10821 | 2269 | 119 |

----------------------------------------------------------------------------

24. Herbert Hoover | March 4 1929 | 19096 | 3799 | 169 |

----------------------------------------------------------------------------

25. James Buchanan | March 4 1857 | 13641 | 2799 | 89 |

----------------------------------------------------------------------------

26. James Garfield | March 4 1881 | 14368 | 2935 | 111 |

----------------------------------------------------------------------------

27. James Madison | March 4 1809 | 5712 | 1167 | 21 |

28. | March 4 1813 | 5813 | 1200 | 32 |

----------------------------------------------------------------------------

29. James Monroe | March 4 1817 | 16101 | 3347 | 121 |

30. | March 4 1821 | 21248 | 4434 | 133 |

----------------------------------------------------------------------------

31. James Polk | March 4 1845 | 23381 | 4763 | 153 |

----------------------------------------------------------------------------

32. Jimmy Carter | January 20 1977 | 5385 | 1210 | 54 |

----------------------------------------------------------------------------

33. John Adams | March 4 1797 | 11223 | 2306 | 37 |

----------------------------------------------------------------------------

34. John F Kennedy | January 20 1961 | 5925 | 1341 | 57 |

----------------------------------------------------------------------------

35. John Quincy Adams | March 4 1825 | 14531 | 2898 | 76 |

----------------------------------------------------------------------------

36. Lyndon B Johnson | January 20 1965 | 6351 | 1498 | 95 |

----------------------------------------------------------------------------

37. Martin Van Buren | March 4 1837 | 19166 | 3807 | 94 |

----------------------------------------------------------------------------

38. Richard Nixon | January 20 1969 | 8883 | 2103 | 107 |

39. | January 20 1973 | 7872 | 1833 | 72 |

----------------------------------------------------------------------------

40. Ronald Reagon | January 20 1981 | 10788 | 2422 | 134 |

41. | January 21 1985 | 11473 | 2560 | 127 |

----------------------------------------------------------------------------

42. Rutherford Hayes | March 5 1877 | 12070 | 2444 | 59 |

----------------------------------------------------------------------------

43. Theodore Roosevelt | March 4 1905 | 4468 | 979 | 33 |

----------------------------------------------------------------------------

44. Thomas Jefferson | March 4 1801 | 8163 | 1711 | 41 |

45. | March 4 1805 | 10465 | 2143 | 45 |

----------------------------------------------------------------------------

46. Ulysses S Grant | March 4 1869 | 5203 | 1107 | 40 |

47. | March 4 1873 | 6170 | 1319 | 43 |

----------------------------------------------------------------------------

48. Warren G Harding | March 4 1921 | 16406 | 3291 | 148 |

----------------------------------------------------------------------------

49. William Clinton | January 20 1993 | 7186 | 1598 | 92 |

50. | January 20 1997 | 9598 | 2153 | 109 |

----------------------------------------------------------------------------

51. William Henry Harrison | March 4 1841 | 40421 | 8401 | 219 |

----------------------------------------------------------------------------

52. William Howard Taft | March 4 1909 | 26134 | 5384 | 160 |

----------------------------------------------------------------------------

53. William McKinley | March 4 1897 | 19187 | 3939 | 130 |

54. | March 4 1901 | 10881 | 2203 | 99 |

----------------------------------------------------------------------------

55. Woodrow Wilson | March 4 1913 | 7631 | 1689 | 67 |

56. | March 4 1917 | 6654 | 1506 | 59 |

----------------------------------------------------------------------------

57. Zachary Taylor | March 5 1849 | 5389 | 1078 | 22 |

OTHER HALF OF TABLE

Speaker |Parags|Sing. | Plur. | Let/Wd| Wd/Snt| Sng/Plr|

----------------------------------------------------------------------------

1. Abraham Lincoln | 1 | 56 | 23 | 4 | 26 | 2.43 |

2. | 5 | 1 | 11 | 4 | 26 | 0.09 |

----------------------------------------------------------------------------

3. Andrew Jackson | 13 | 32 | 21 | 4 | 44 | 1.52 |

4. | 10 | 25 | 28 | 4 | 38 | 0.89 |

----------------------------------------------------------------------------

5. Barack Obama | 1 | 5 | 147 | 4 | 19 | 0.03 |

6. | 30 | 4 | 153 | 4 | 22 | 0.03 |

----------------------------------------------------------------------------

7. Benjamin Harrison | 36 | 19 | 108 | 4 | 27 | 0.18 |

----------------------------------------------------------------------------

8. Calvin Coolidge | 24 | 10 | 150 | 4 | 20 | 0.07 |

----------------------------------------------------------------------------

9. Dwight Eisenhower | 53 | 9 | 123 | 4 | 20 | 0.07 |

10. | 42 | 4 | 74 | 4 | 17 | 0.05 |

----------------------------------------------------------------------------

11. Franklin Pierce | 15 | 59 | 55 | 4 | 31 | 1.07 |

----------------------------------------------------------------------------

12. Franklin Roosevelt | 37 | 12 | 85 | 4 | 18 | 0.14 |

13. | 34 | 2 | 47 | 4 | 17 | 0.04 |

14. | 17 | 6 | 33 | 4 | 18 | 0.18 |

15. | 25 | 29 | 58 | 4 | 21 | 0.50 |

----------------------------------------------------------------------------

16. George Bush | 28 | 35 | 111 | 4 | 15 | 0.32 |

----------------------------------------------------------------------------

17. George W Bush | 30 | 12 | 99 | 4 | 16 | 0.12 |

18. | 28 | 9 | 88 | 4 | 20 | 0.10 |

----------------------------------------------------------------------------

19. George Washington | 7 | 51 | 3 | 4 | 61 | 17.00 |

20. | 3 | 8 | 0 | 4 | 33 | inf |

----------------------------------------------------------------------------

21. Grover Cleveland | 19 | 11 | 34 | 4 | 37 | 0.32 |

----------------------------------------------------------------------------

22. Grover cleveland | 27 | 30 | 74 | 5 | 34 | 0.41 |

----------------------------------------------------------------------------

23. Harry Truman | 73 | 7 | 80 | 4 | 19 | 0.09 |

----------------------------------------------------------------------------

24. Herbert Hoover | 50 | 24 | 121 | 5 | 22 | 0.20 |

----------------------------------------------------------------------------

25. James Buchanan | 22 | 23 | 49 | 4 | 31 | 0.47 |

----------------------------------------------------------------------------

26. James Garfield | 39 | 16 | 49 | 4 | 26 | 0.33 |

----------------------------------------------------------------------------

27. James Madison | 7 | 30 | 14 | 4 | 55 | 2.14 |

28. | 11 | 9 | 27 | 4 | 37 | 0.33 |

----------------------------------------------------------------------------

29. James Monroe | 21 | 38 | 101 | 4 | 27 | 0.38 |

30. | 28 | 45 | 80 | 4 | 33 | 0.56 |

----------------------------------------------------------------------------

31. James Polk | 31 | 47 | 111 | 4 | 31 | 0.42 |

----------------------------------------------------------------------------

32. Jimmy Carter | 27 | 17 | 77 | 4 | 22 | 0.22 |

----------------------------------------------------------------------------

33. John Adams | 14 | 32 | 10 | 4 | 62 | 3.20 |

----------------------------------------------------------------------------

34. John F Kennedy | 28 | 5 | 59 | 4 | 23 | 0.08 |

----------------------------------------------------------------------------

35. John Quincy Adams | 10 | 35 | 48 | 5 | 38 | 0.73 |

----------------------------------------------------------------------------

36. Lyndon B Johnson | 35 | 22 | 76 | 4 | 15 | 0.29 |

----------------------------------------------------------------------------

37. Martin Van Buren | 20 | 82 | 101 | 5 | 40 | 0.81 |

----------------------------------------------------------------------------

38. Richard Nixon | 71 | 20 | 122 | 4 | 19 | 0.16 |

39. | 52 | 16 | 99 | 4 | 25 | 0.16 |

----------------------------------------------------------------------------

40. Ronald Reagon | 36 | 25 | 125 | 4 | 18 | 0.20 |

41. | 41 | 19 | 127 | 4 | 20 | 0.15 |

----------------------------------------------------------------------------

42. Rutherford Hayes | 28 | 35 | 26 | 4 | 41 | 1.35 |

----------------------------------------------------------------------------

43. Theodore Roosevelt | 5 | 1 | 62 | 4 | 29 | 0.02 |

----------------------------------------------------------------------------

44. Thomas Jefferson | 7 | 31 | 42 | 4 | 41 | 0.74 |

45. | 15 | 33 | 44 | 4 | 47 | 0.75 |

----------------------------------------------------------------------------

46. Ulysses S Grant | 15 | 29 | 15 | 4 | 27 | 1.93 |

47. | 17 | 47 | 16 | 4 | 30 | 2.94 |

----------------------------------------------------------------------------

48. Warren G Harding | 33 | 20 | 155 | 4 | 22 | 0.13 |

----------------------------------------------------------------------------

49. William Clinton | 17 | 13 | 117 | 4 | 17 | 0.11 |

50. | 24 | 7 | 124 | 4 | 19 | 0.06 |

----------------------------------------------------------------------------

51. William Henry Harrison | 25 | 87 | 85 | 4 | 38 | 1.02 |

----------------------------------------------------------------------------

52. William Howard Taft | 44 | 49 | 71 | 4 | 33 | 0.69 |

----------------------------------------------------------------------------

53. William McKinley | 19 | 38 | 86 | 4 | 30 | 0.44 |

54. | 13 | 20 | 63 | 4 | 22 | 0.32 |

----------------------------------------------------------------------------

55. Woodrow Wilson | 11 | 5 | 71 | 4 | 25 | 0.07 |

56. | 17 | 14 | 90 | 4 | 25 | 0.16 |

----------------------------------------------------------------------------

57. Zachary Taylor | 10 | 32 | 22 | 4 | 49 | 1.45 |

Code:

/\*

\* Speech analyzer

\*

\* Class: CS 141, Spring 2014, Thurs 4pm lab.

\* System: Linux Mint x64, g++

\* Author: Brian Howe

\*

\* TODO: output is off for some speeches and singular pronouns.

\*/

#include <stdio.h>

#include <stdlib.h>

#include <string.h>

#include <ctype.h>

struct speech\_struct {

char speaker[32];

char date[32];

int letters;

int words;

int sents;

int parags;

int sing;

int plur;

} typedef SPEECH;

char \*sing[7] = {(char\*)"i", (char\*)"i've", (char\*)"i'll", (char\*)"me",

(char\*)"my", (char\*)"mine", (char\*)"myself"};

char \*plur[7] = {(char\*)"we", (char\*)"we've", (char\*)"we'll", (char\*)"us",

(char\*)"our", (char\*)"ours", (char\*)"ourselves"};

void print\_heading(int num, char \*s) {

printf("Author: Brian Howe\n");

printf("Lab: Thur 4pm\n");

printf("Program: #%d, %s\n\n", num, s);

}

// changes teh size of a char array

char \* resize(char \*in, int n) {

free(in);

char \*out = (char\*)malloc(sizeof(char) \* n);

return out;

}

// finds all singular and plural pronouns

void pronouns(SPEECH \*sData, char \*word) {

int i;

for(i = 0; i < 7; i++) {

if(strcmp(word, sing[i]) == 0) {

sData->sing++;

}

else if(strcmp(word, plur[i]) == 0) {

sData->plur++;

}

}

}

// analyzes all the content of the file (letters, words, paragaphs, etc.)

void get\_data(SPEECH \*sData, FILE \*sFile) {

char \*all;

// initialize values

sData->letters = 0;

sData->words = 0;

sData->sents = 0;

sData->parags = 1; // the last parag doesnt end in '\r'

sData->sing = 0;

sData->plur = 0;

int n = 20000; // 20000 is an average amount of characters in the speeches

all = (char\*)malloc(sizeof(char) \* n);

do {

fseek(sFile, SEEK\_SET, 0); // go to beginning of file

fread(all, sizeof(char), n, sFile); // read file to array 'all'

if(!feof(sFile)) { // if all was to small to fit the file

n+=1000; // increase size of all

all = resize(all, n);

}

} while(!feof(sFile));

char \*word = strtok(all, " "); // read first word in file

int lenWord, i;

while(word != NULL) { // while there are still words in fill

sData->words++;

lenWord = strlen(word);

for(i = 0; i < lenWord; i++) {

if(isalpha(word[i])) {

sData->letters++;

word[i] = tolower(word[i]);

}

else if(word[i] == '.' || word[i] == '?') {

sData->sents++;

}

else if(word[i] == '\r') {

sData->parags++;

i = lenWord; // avoids extra '\r'

}

}

pronouns(sData, word);

word = strtok(NULL, " "); // read next word

}

}

// opens file and parses name. then calls get\_data

void analyze(SPEECH \*sData, char \*path) {

FILE \*sFile; // the speech file

sFile = fopen(path, "r"); // opens the speech

if(sFile == NULL) {

printf("Could not open %s.\n", path);

}

path = &path[10]; // removes datafiles/ from string

char \*tmp = path;

while(tmp != NULL) { // changes all '\_' to ' '

tmp = strchr(path, '\_');

if(tmp != NULL) {

\*tmp = ' ';

}

}

char \*tmpN = strtok(path, "-"); // gets the speakers name

char \*tmpD = strtok(NULL, "."); // gets the date of speech

strcpy(sData->speaker, tmpN);

strcpy(sData->date, tmpD);

get\_data(sData, sFile);

}

// prints a char a specified amount

void print\_char(char c, int n) {

int i;

for(i = 0; i <= n; i++) {

printf("%c", c);

}

}

// prints all speech data

void print\_data(SPEECH s[], int n) {

int i;

printf("%26s | %15s | %6s | %6s | %6s | %6s | %6s | %6s | %6s| %6s| %7s|\n",

"Speaker",

"Date",

"Letrs",

"Words",

"Sents",

"Parags",

"Sing.",

"Plur.",

"Let/Wd",

"Wd/Snt",

"Sng/Plr");

for(i = 0; i < n; i++) {

if(i > 0 && strcmp(s[i-1].speaker, s[i].speaker) == 0) {

printf("%2d.", i+1);

print\_char(' ', 23);

printf("| ");

}

else{

print\_char('-', 75);

printf("\n%2d.%23s | ", i+1, s[i].speaker);

}

printf("%15s | ", s[i].date);

printf("%6d | ", s[i].letters);

printf("%6d | ", s[i].words);

printf("%6d | ", s[i].sents);

printf("%6d | ", s[i].parags);

printf("%6d | ", s[i].sing);

printf("%6d |", s[i].plur);

printf("%6d |", s[i].letters/s[i].words);

printf("%6d |", s[i].words/s[i].sents);

printf("%7.2f |", (double)s[i].sing/s[i].plur);

printf("\n");

}

}

int main() {

FILE \*speechList;

SPEECH speeches[56];

print\_heading(4, (char\*)"Inaugural Analysis");

speechList = fopen("datafiles.txt", "r");

if(speechList == NULL) {

printf("Could not open file speeches.txt.\n");

return -1;

}

int i = 0;

char \*tmp;

char \*buff = (char\*) malloc(sizeof(char) \* 128);

while(!feof(speechList) && i < 57) {

strcpy(buff, (char\*)"datafiles/");

fgets(&buff[10], 119, speechList);

tmp = strchr(buff, '\n');

if(tmp != NULL) {

\*tmp = '\0';

}

analyze(&speeches[i], buff);

i++;

}

print\_data(speeches, i);

return 0;

}