Name: Jameel Kaba
UWO/GAUL id: jkaba
Student #: 250796017

**CS4442B Assignment 3**

**Question 1**

**A)**

Refer to P1.cpp

**B)**

Dostoevsky Karamazov Text

You need k = 60 in order to get close to 50%.

K = 60 gives you 50.0261% Output below

the, 15173

and, 11436

to, 9528

he, 8149

i, 7995

of, 7281

a, 6804

you, 6305

that, 6177

it, 5852

in, 5514

was, 4776

his, 4345

s, 3480

for, 3471

him, 3377

but, 3267

at, 3004

with, 2986

not, 2907

had, 2854

is, 2433

on, 2334

me, 2292

all, 2196

as, 2132

t, 2108

have, 1983

her, 1885

be, 1874

she, 1784

what, 1743

so, 1705

my, 1691

one, 1603

from, 1486

there, 1357

they, 1310

alyosha, 1243

this, 1241

are, 1222

by, 1205

no, 1205

will, 1168

if, 1162

been, 1105

would, 1102

up, 1091

your, 1064

only, 1055

were, 1000

said, 995

them, 987

out, 959

an, 943

now, 927

mitya, 917

man, 908

who, 898

do, 866

50.0261 %


* Runs Dr. Seuss Text fast *

Dr. Seuss Text

You need k = 36 to get just above 50% (50.5176%)

the, 89
and, 66
i, 62
of, 45
a, 41
king, 32
that, 31
he, 23
to, 22
all, 19
in, 19
yertle, 19
turtle, 18
m, 16
one, 15

turtles, 15
you, 15
my, 12
s, 12
they, 12
them, 11
up, 11
was, 11
going, 10
his, 10
it, 10
throne, 10
but, 9
on, 9
see, 9
t, 9
down, 8
here, 8
little, 8
ll, 8
me, 8
50.5176 %

## Question 2

**A)**

Refer to P2.cpp for code

**B)**

For N Grams of size 1: 33.1106%

For N Grams of size 2: 68.6534%

For N Grams of size 3: 87.5173%

For N Grams of size 4: 96.7292%

For N Grams of size 5: 99.2807%

For N Grams of size 6: 99.837%

For N Grams of size 7: 99.9504%

For N Grams of size 8: 99.9805%

For N Grams of size 9: 99.9889%

For N Grams of size 10: 99.9922%

For N Grams of size 11: 99.995%

For N Grams of size 12: 99.9961%

For N Grams of size 13: 99.9972%

For N Grams of size 14: 99.9978%

For N Grams of size 15: 99.9983%

For N Grams of size 16: 99.9989%

For N Grams of size 17: 99.9994%

For N Grams of size 18: 100%

Longest N Gram is N = 17: repulsion that s what i m afraid of that s what may be too much for me

Value of N which gives no common N grams: 18

**C)**

For N Grams of size 1: 32.8801%

For N Grams of size 2: 77.4582%

For N Grams of size 3: 94.5332%

For N Grams of size 4: 99.1954%

For N Grams of size 5: 99.8854%

For N Grams of size 6: 99.9768%

For N Grams of size 7: 99.9977%

For N Grams of size 8: 100%

Longest N Gram is N = 7: there is no such thing as a

Longest N Gram is N = 7: in the middle of the table and

Value of N which gives no common N grams: 8


**D)**

For N Grams of size 1: 84.1851%

For N Grams of size 2: 97.468%

For N Grams of size 3: 99.5337%

For N Grams of size 4: 99.9199%

For N Grams of size 5: 99.987%

For N Grams of size 6: 99.9984%

For N Grams of size 7: 100%

Longest N Gram is N = 6: from them what they have not

Longest N Gram is N = 6: of a man s own labour

Longest N Gram is N = 6: in order to keep up the

Longest N Gram is N = 6: to keep up the rate of

Longest N Gram is N = 6: is the same as that of

Longest N Gram is N = 6: of nature and of reason the

Value of N which gives no common N grams: 7

**E)**

The comparison of the 2 parts of Dostoevsky result in the longest length of N Gram. I think that this could be in part since Dostoevsky wrote both parts, which would result in him using similar words in both parts.

With the other comparisons, since both texts which were compared (in both c and d) were written by different authors, the likelihood that they would use similar phrases is unlikely. As such, the length of N Grams is smaller for both of these.

Interestingly, in the comparison of Marx and Smith (part d), the percentage is quite high when N = 1, compared to the others. It may be the case in which there are similar phrases in both texts, but they could just be worded differently which leads it to not being an N Gram.

**Question 3**

Broken Code / Not finished

Refer to P3.cpp

**3B, 3C, 3D**

Program not complete, will not run

**Question 4**

Broken Code / Not finished

Refer to P4.cpp

**4B)**

Code returns 0 for all options, needs further work to be completed

**Question 5**

Broken Code / Not finished

Refer to P5.cpp

**5B)**

Code returns inf for all options, needs further work to be completed

**Question 6**

Did not attempt

**Question 7**

Did not attempt