

An Exploration of Metamorphic Malware

Research Question: **What methods of Virial metamorphism are most effective at creating unique viral families as measured through signature based detection?**

Personal Code: jkz289

International Baccalaureate Extended Essay

Computer Science

May 2021

Word Count: 3992

1.Table of Contents

1.Table of Contents	2
2. Introduction	4
2.1 Viruses	4
2.2 Metamorphic Virus	5
2.2.1 Garbage Code Insertion	6
2.2.2 Permutation	6
2.2.3 Code Integration	6
3. Metamorphic Virus Payloads	7
4. Methodology	7
4.1 Creating Virus Strings and File Strings	8
4.1.1 Garbage Code Insertion Implementation (GCI)	8
4.1.2 Permutation Implementation	9
4.1.3 Code Integration Implementation	9
4.2 Creation of Virus-Files	9
4.3 Comparing Two Virus-Files	10
5. Data Collection	11
5.1 Sample Graph without Noise removal	11
5.2 Sample Scatter Plots	12
5.3 Scatter Plot Data Observations	13
5.4 Heat Map Creation	14
5.5 Enlarged Heat Map Creation	16
5.6 Heat Map Observations	17
6. Quantizing Data	17
6.1 Determining Strength of Virus	17
6.2 Strength of Virus Table	18
6.3 Calculating Strength of Metamorphism	19
6.4 Strength of Metamorphism Table	19
7. Analysis	19
7.1 Simple Virus and Integration Virus	19
7.2 Permutation Virus	20
7.3 GCI Virus	21
8. Conclusion	22

9. Bibliography	23
10. Appendix	25
10.1. Java Code	25
10.1.1. Garbage Code Insertion Algorithm	25
10.1.2. Permutation Algorithm	27
10.1.3. Integration Algorithm	28
10.1.4. Compare Algorithm	29
10.2. Texts	35
10.2.1. Virus String	35
10.2.2. Essay 3	36
10.2.3. Essay 4	38
10.2.4. Essay 5	40
10.2.5. Essay 6	43
10.2.6. Essay 7	45
10.3. Graphs	49
10.3.1. Scatter Plots	49
10.3.2. Heat Map	53
10.3.3. Enlarged Heat Map	56

2. Introduction

This essay will investigate the methods by which viruses attempt to avoid static, signature-based detection. That is, detection that is based on the analysis of a single stagnant source of information. Specifically, this paper will analyze the common techniques that metamorphic viruses use to avoid detection and a determination on which of these techniques are most effective at avoiding detection will be made.

2.1 Viruses

A virus is a computer program that delivers malicious software to computers. This program can damage files, give others access to private information, or even affect the computer's speed. Much like biological viruses, computer viruses spread by inserting copies of itself into other files in the computer. As the virus spreads, the virus code does not change.¹

Simple Virus Evolution



Figure 2.1.1: The unchanging boxes represent the unchanging body of the virus code.

This virus is vulnerable to anti-virus programs that check files for virus signatures.¹ These virus signatures are like a virus's unique fingerprint that each copy of the virus keeps as it spreads. An anti-virus using this method of detection would have an entire database of known virus fingerprints. By checking the file for these fingerprints in its database, anti-virus software can

¹ Konstantinou, Evgenios. *Metamorphic Virus: Analysis and Detection*. 15 Jan. 2008, <https://www.ma.rhul.ac.uk/static/techrep/2008/RHUL-MA-2008-02.pdf>. Accessed 10 Feb. 2021.

identify and remove files that have known fingerprints. This form of detection is called static, signature based detection.

2.2 Metamorphic Virus

Metamorphic viruses were created as a result of anti-virus software becoming increasingly effective at detecting viruses. Each generation of a metamorphic virus creates a unique fingerprint. Since the fingerprint is unique, the anti-virus software does not have the new signature in its database and will not know that the file contains a virus. This renders the static signature based detection useless. Metamorphic viruses accomplish this feat by rewriting their code from generation to generation. This gives the virus the ability to completely change the size, location, and even the contents of virus code.²

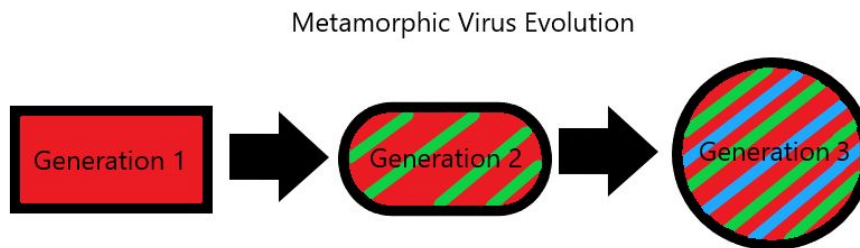


Figure 2.1.2: The changing shapes represent the changing body of the virus code.

Even though the code of the virus changes, through clever programming, the virus is able to maintain basic functions like continue infecting other files or do damage to infected files.

Metamorphic viruses utilize a variety of techniques to re-write their code. Here are some of these techniques:

² Szor, Peter and Peter Ferrie. *Hunting For Metamorphic*. Symantec, <https://crypto.stanford.edu/cs155old/cs155-spring09/papers/viruses.pdf>. Accessed 15 Feb. 2021. Accessed 15 Feb. 2021.

2.2.1 Garbage Code Insertion

Viruses using garbage code insertion will add do-nothing code to itself. This code, as the name implies, does not accomplish anything and exists solely to change the appearance of the virus. Each time the virus spreads to a new file, it will implement different amounts of do-nothing code making each generation of the virus look unique. This technique is utilized in the Win32/Evol and Win95/Bistro viruses. ³

2.2.2 Permutation

Metamorphic viruses using the permutation technique will first divide its virus code into many different segments. These segments will be swapped around every time a virus spreads to a new file creating viruses that look unique. This technique is used by Win32/Ghost and Win95/Zmist.

2.2.3 Code Integration

Viruses using code integration take an entire file, decompartmentalizes it, then places itself in some random spot inside the file, and then reassembles the whole file. Everytime the virus spreads, it will place itself in a new spot in the infected file. This can be a very effective technique since most antivirus software only looks at specific locations where there are likely to be viruses. But, if the virus changes location every time it spreads, then this method of searching for viruses does not work. Apart from using permutation, the Win95/Zmist Virus also makes use of code integration. ⁴

³ Konstantinou, Evgenios. *Metamorphic Virus: Analysis and Detection*.

⁴ Beaucampshilippe Beaucamps. Advanced Metamorphic Techniques in Computer Viruses. International Conference on Computer, Electrical, and Systems Science, and Engineering - CESSE'07, Nov 2007, Venice, Italy. ffinria-00338066f

3. Metamorphic Virus Payloads

The part of the virus that does damage to the computer is called its payload. This payload in malware like MyDoom, a computer worm that infected over 500,000 machines, has caused 38 billion dollars worth of damage.⁵ Thankfully, metamorphic viruses are very difficult to create which is why there are not many of them. However, these viruses are so elusive that if a computer were to get infected, it is unlikely that the user would ever know unless the virus were to release its payload. In order to detect metamorphic viruses before they cause extreme economic damage, each metamorphic technique must be studied in depth. Furthermore, because of how dangerous metamorphic viruses are, real metamorphic viruses cannot be used and a different method must be used to analyze them.

4. Methodology

To understand the strengths and weaknesses of each metamorphic technique, a string representation of a virus and a file will be created. This virus string will undergo a metamorphic transformation to create a new generation of virus. This new generation of virus will subsequently be placed into a new file string. The newly created virus-file string will be compared to the other virus-files created with the same metamorphic techniques and a determination will be made for what the impact of the metamorphic transformation was to each generation of virus-files.

⁵ Duncan, Brad. "MyDoom Still Active in 2019." *paloalto*, 26 July 2019, <https://unit42.paloaltonetworks.com/mydoom-still-active-in-2019/>. Accessed 2 Mar. 2021.

4.1 Creating Virus Strings and File Strings

When just analyzing virial metamorphism, a copy of a real metamorphic virus is not needed.

Instead, the parts of the virus that allow it to transform are needed. These parts can be recreated by making transformation algorithms in Java. Additionally, the algorithm does not have to transform actual virus machine code and can instead transform a string that represents a virus.

Furthermore, a real computer file is not needed to examine the effects of a file being infected by a virus. Instead, a large string that represents a file can be “infected”, by placing the transformed metamorphic virus string inside the file string. After some preliminary data collection, I found that substituting virus machine code and real computer files with string representations of them does not impact the virus’s ability to transform and infect files since the results produced were similar to that produced in other metamorphic virus research papers.⁶ Here are three different metamorphism algorithm created with Java that will be used to analyze metamorphic virus transformation:

4.1.1 Garbage Code Insertion Implementation (GCI)

To replicate this technique, I will utilize an array list that will contain strings of “garbage” text. Each time the algorithm is implemented, a random amount of strings from the garbage text array list will be selected, combined into one big string, and then inserted into a random position in the virus string. This algorithm will be implemented a random amount of times in one generation of the metamorphic virus. For example, if the Virus String is “I am a Virus” and the garbage text arraylist contains {“Garbage”, “Trash”, “Litter”}, then the resulting GCI string might look like “I

⁶ Wong, Wing and Mark Stamp. *Hunting for Metamorphic Engines*. Springer, 11 Nov. 2006, <https://ptolemy.berkeley.edu/projects/truststc/pubs/237/hunting.pdf>. Accessed 16 Feb. 2021.

GarbageLitter am a LitterTrashGarbage Virus”. The code for the GCI algorithm is located in the appendix section 10.1.1.

4.1.2 Permutation Implementation

For this technique, the algorithm will take each word in the virus and relocate it in a different random position of the virus. So, if the virus string is “I am a Virus”, then the permutation virus string might look like “a I Virus am”. The code for the permutation algorithm is located in the appendix section 10.1.2.

4.1.3 Code Integration Implementation

In the previous algorithms, the virus is positioned in the same location in the file string. However, the code integration algorithm will randomize the position of the virus string in the file string. So, if the file string contains 50 words, maybe one generation of the virus string will be placed 10 words from the start of the file string while the next generation would be placed 30 words from the start of the file string. The code for the Integration algorithm is located in the appendix section 10.1.3.

4.2 Creation of Virus-Files

Once the virus string has been transformed, the virus can then be placed inside the file string. These file strings will be large essays that follow proper grammar and linguistic rules much like how real computer files follow certain formatting rules. These essays will be larger than the virus, since most viruses will not be larger than the computer file they are infecting. They will also contain information about a variety of different topics and have different word counts because real computer files come in a variety of different sizes and contain different information

depending on the file. The essays will be taken from an essay database called OEssay.⁷ After the file strings are chosen, the virus string can be inserted into them. This new virus string, file string amalgamation is called a virus-file which will be used for analysis.

4.3 Comparing Two Virus-Files

To understand the effects of the metamorphic algorithms, two generations virus-files must be compared in order to find the differences between the two generations. Understanding the differences between each generation of virus-files is important for determining the range of transformations the algorithm is capable of doing. This comparison can be done using an adaptation to Puneet Mishra's Opcode similarity test as seen in Taxonomy of Uniqueness Transformations.⁸ In this adaptation, each of the two virus-file strings are divided into smaller strings called substring. Each substring will contain 100 sentences from the virus-file. These substrings will be added to one of two arraylists depending on whether the sentences are from one virus-file or the other. From there, the arraylists will compare the contents of the substrings that are in the same index of the arraylist. So, the first 100 sentences of one virus-file is compared to the first 100 sentences in the other virus-file, the next 100 sentences in one virus-file is compared to the next 100 sentences in the other virus-file, etc. If two words in the substrings from each arraylist are exactly the same, then the position of the first and second words become x-y coordinates for a point that will be plotted in a scatter plot. For example, if in the first 100 sentences of one virus-file, the word "Hello." is located, and in the first 100 sentences of the other virus-file, the same word is found, then, the program will count how many

⁷ "Oessays - The Biggest Essays Database of 2021.". oessays, <https://oessays.com/>. Accessed 16 Feb. 2021. Accessed 16 Feb. 2021.

⁸ Mishra, Puneet. *Taxonomy of Uniqueness Transformations*. San Jose State University, Dec. 2003. Accessed 16 Feb. 2021.

characters are between the start of each virus-file and the word “Hello.” Those numbers will become the x-y coordinates. The entire file is scanned in this way creating the full scatter plot. The code that compares the virus-files is located in section 10.1.4.

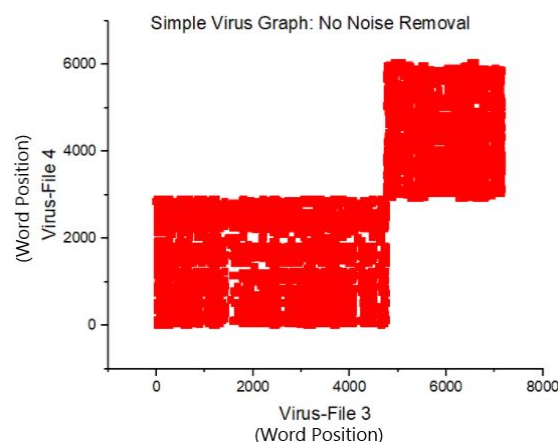
5. Data Collection

The data that is going to be collected will be produced from a single unchanging virus string that is 10 sentences long and can be found in the appendix section 10.2.1. The virus will be implemented in one of five ways:

- 1) No virus is inserted (Control)
- 2) No transformation inserted at sentence N (Non-metamorphic virus)
- 3) Use GCI algorithm inserted at sentence N
- 4) Use permutation algorithm inserted at sentence N
- 5) Using integration algorithm which places virus string in random locations

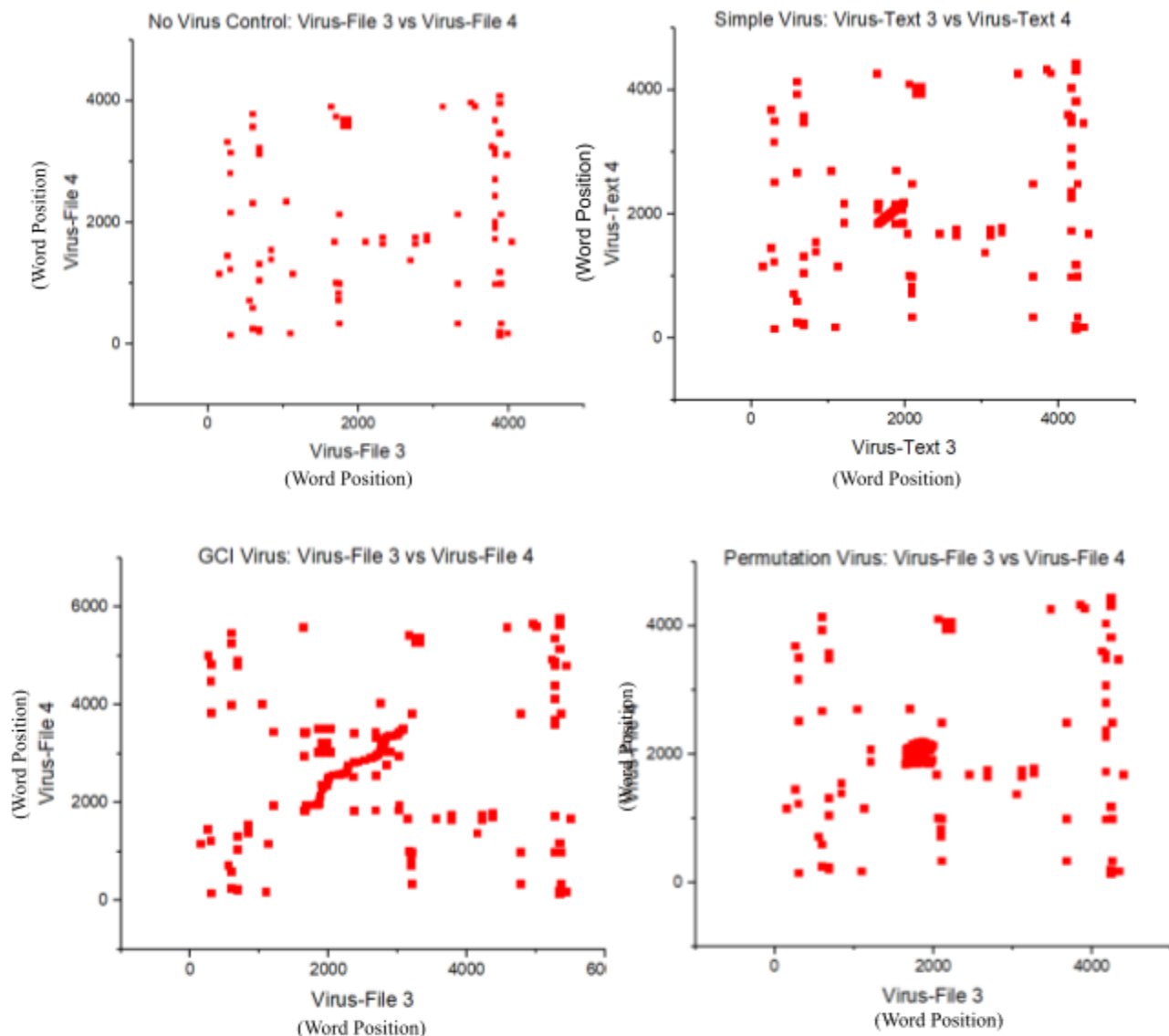
Five essays from OEssay will be used to create five file strings which can be found in the appendix section 10.2.2-10.2.6. Virus-files created using the same metamorphic technique will be compared. One of the virus-files (virus-file 3) will be compared to other virus-files (virus-files 4-7). Four scatter plots will be produced for each of the five implementations, so 20 scatter plots in total will be produced.

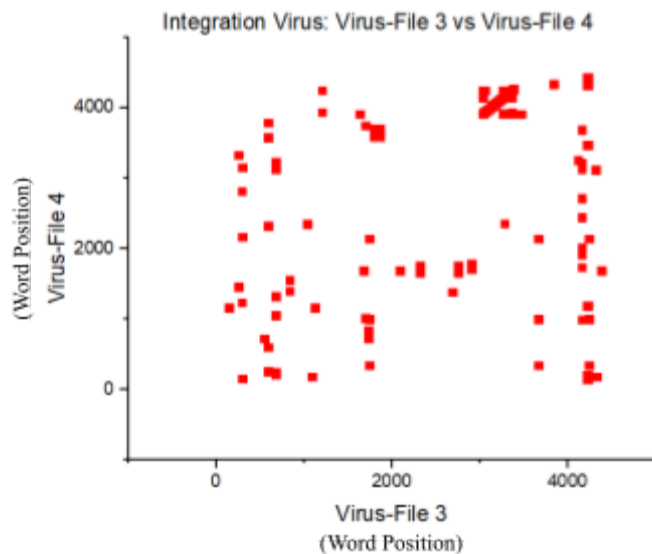
5.1 Sample Graph without Noise removal



This is a sample virus-file comparison scatter plot. At the moment, common English words such as “a”, “have”, “their”, etc are producing points in such great numbers that it obscures any pattern that could have been produced by the virus string. To more easily see these patterns, an algorithm that eliminates the most common English words will be used to remove the noise.

5.2 Sample Scatter Plots





The rest of the scatter plots are located in the appendix section 10.3.1.

5.3 Scatter Plot Data Observations

In the control scatter plot, points were arranged in random positions with no correlation. There were some areas where vertical lines of points were created such as at virus-file 3 word position 4000 but these are just common words that have not been filtered out. Recurring points from the control scatter plot in the other graphs can be disregarded since they were not created by the virus string. In the scatter plots for the simple and integration viruses, there is a set of points that are closely correlated that form a diagonal line. These diagonals in the integration virus scatter plot are located in random areas of the graph. This is different from the simple virus whose diagonals were consistently located in the center of the graph. For the GCI virus, there is a correlation of points in the center of the graph that form a rough, elongated diagonal line. In the permutation virus, there is also correlation around the center but, instead of a diagonal line, a rectangular area of points is located there.

5.4 Heat Map Creation

In order to remove the points that are not created by the virus string, I will convert the scatter plots into heatmaps. Heatmaps illustrate the places and the magnitudes of concentrations of points. They accomplish this by creating vertical and horizontal lines in the scatter plots creating a matrix of boxes.

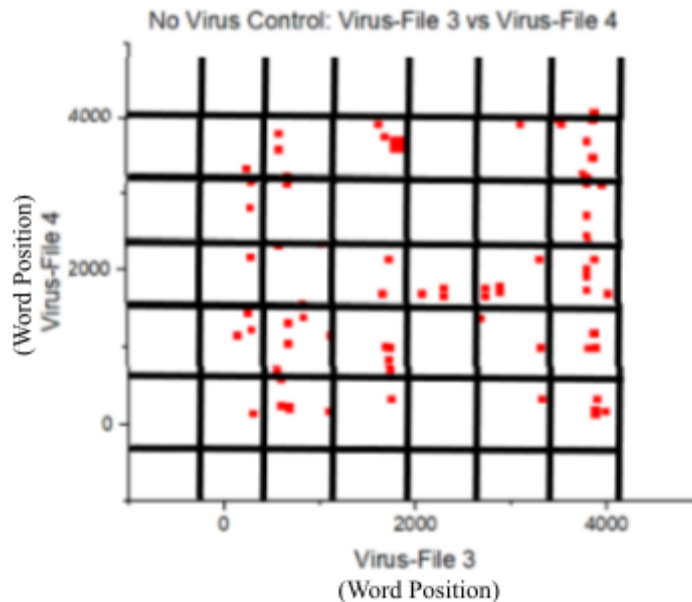
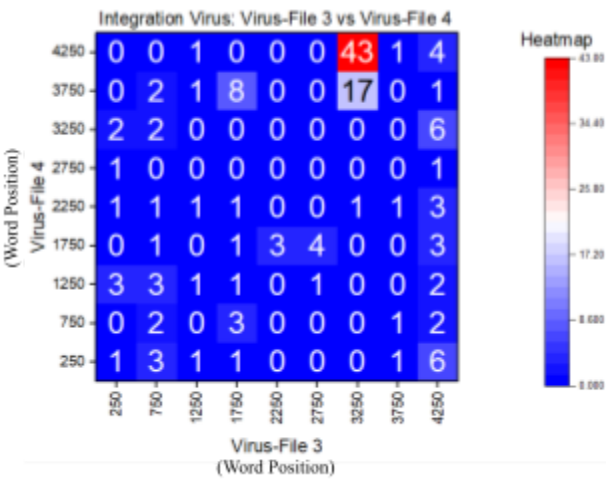
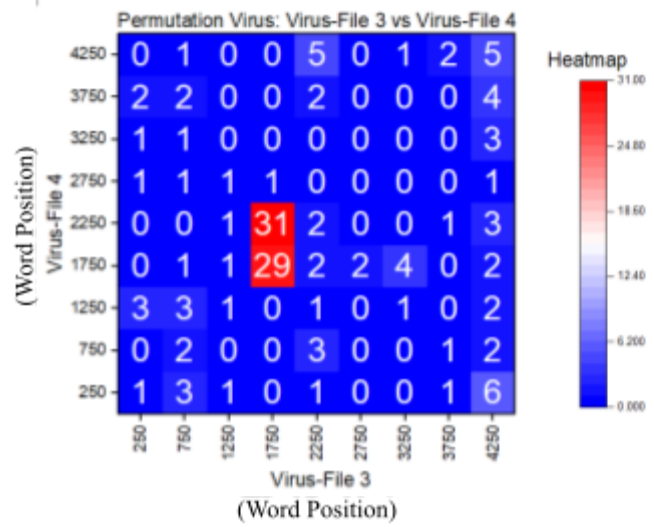
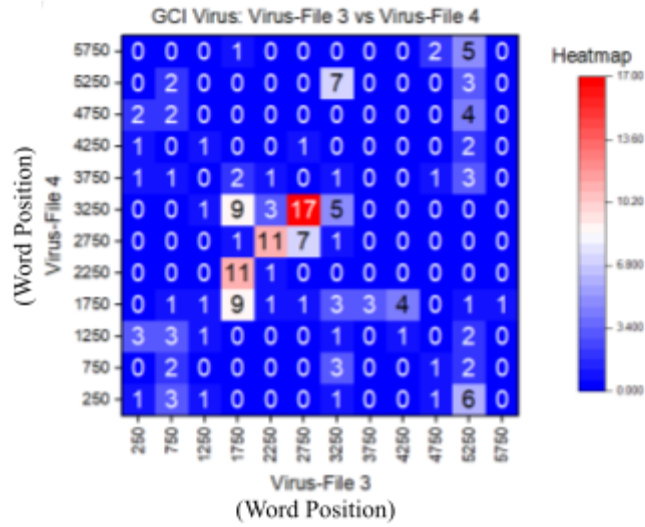
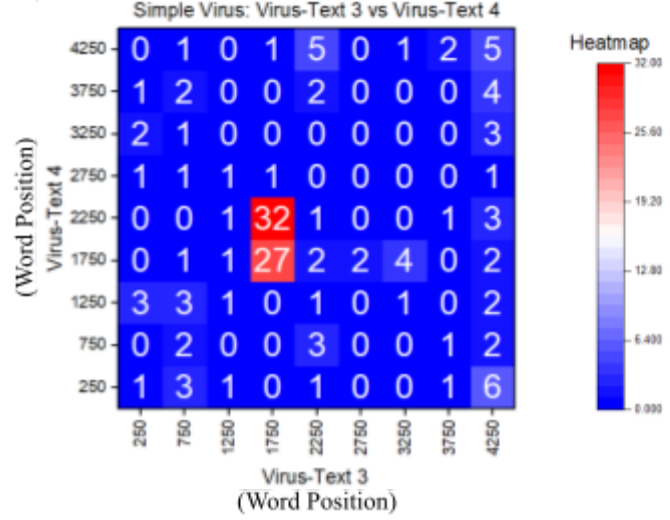
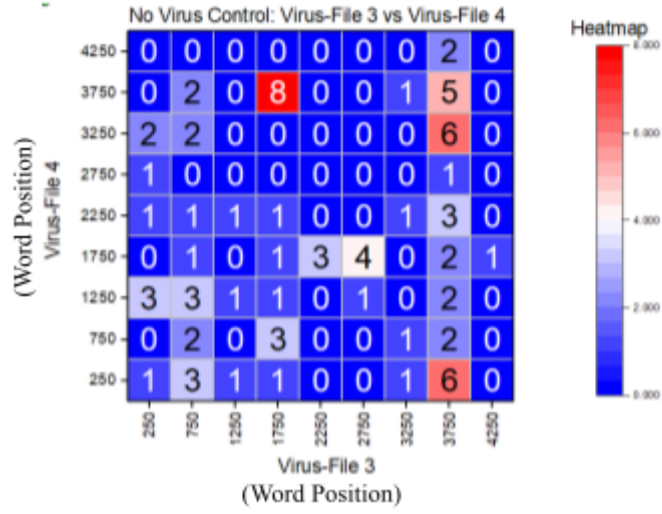


Figure 5.4.1: Matrix of boxes in scatter plot

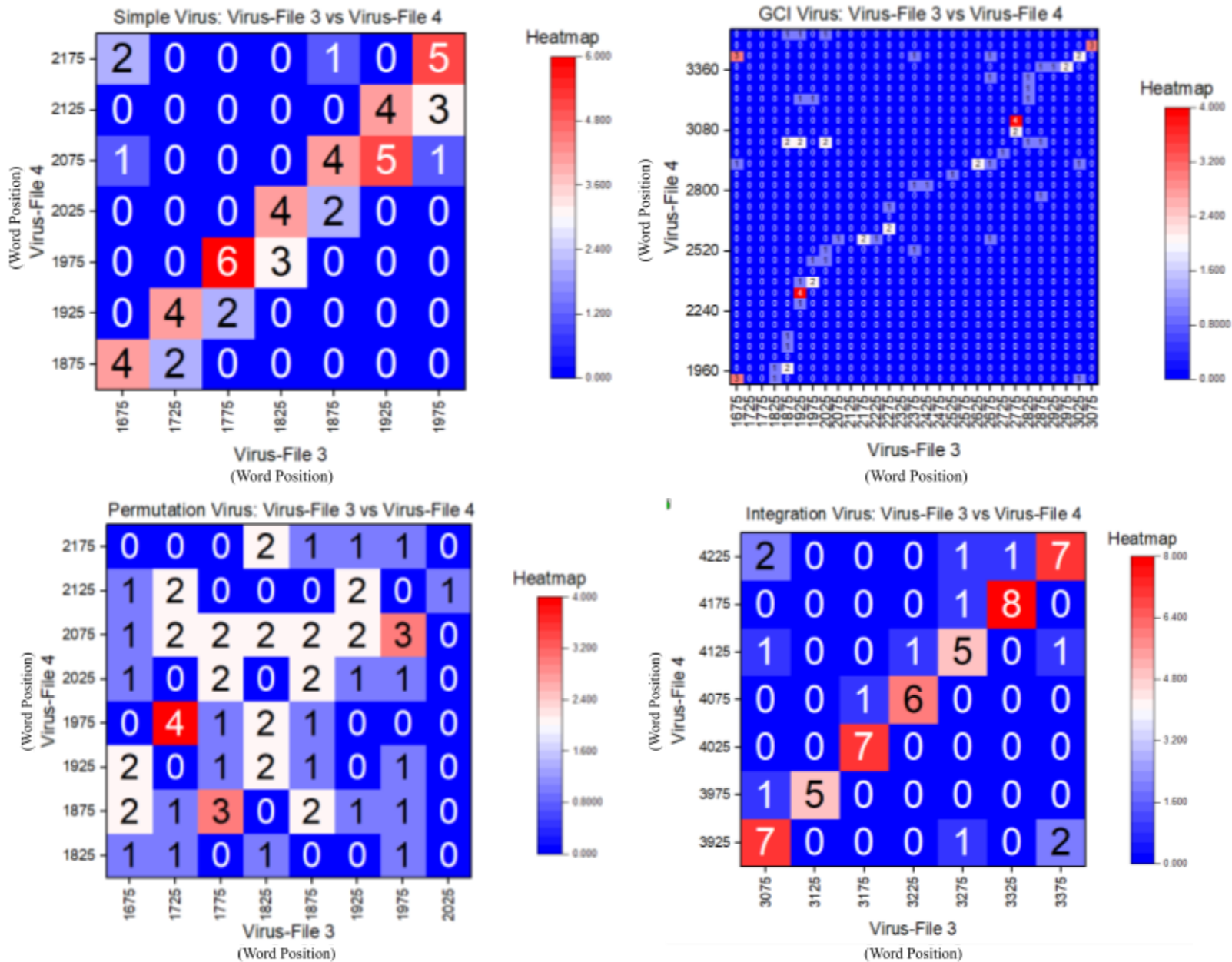
The boxes represent a subsection of the domain and range of the graph and the total amount of points in each box is then displayed in their respective boxes. Each box is then colored based on the values in the boxes. Large values are colored with warm colors while low values are colored with cooler colors. The lengths and widths of the boxes control the amount of boxes in the graph and represent the resolution of the heatmap. Boxes with high values indicate that there are many points in that domain-range subsection which means that there is likely a virus located in that area. I am trying to encapsulate every single virus point in one box while minimizing the amount of extra boxes that I will have to create. After some experimentation with the lengths and widths, I determined that the best value for both is 500.



The rest of the heat maps are located in the appendix section 10.3.2.

5.5 Enlarged Heat Map Creation

Now that the locations of the viruses have been determined, a much finer resolution heat map of the virus can be produced. These heatmaps will have boxes that have lengths and widths equal to 50 which will help reveal information about the concentrations of points within the virus.



The rest of the enlarged heat maps are located in the appendix section 10.3.3.

5.6 Heat Map Observations

In both the simple and integration virus heat maps, the maps show high box values along the diagonal. For the permutation virus, the values are evenly distributed across the heatmap. The GCI virus is distributed amongst many more boxes in the regular heat map so it will have many boxes in the enlarged heat map. The GCI virus in the enlarged heat map forms a semi-straight line with oftentimes gaps of boxes with low or zero value between boxes with higher values.

6. Quantizing Data

To determine the effectiveness of the transformation algorithms, the strength of metamorphism for each transformation must be determined. To do this, I will quantify how different the virus is between two generations called the strength of the virus and will then determine the statistical variations between the virus strengths which will give me the strength of metamorphism.

6.1 Determining Strength of Virus

In order to determine the strength of the virus, I am going to find its viral diffusion and viral potency. Viral diffusion represents the change in the spread of the virus across each of the virus-files. What changes to the size of the virus string did the algorithm make? If there are a lot of boxes with values in them in comparison to the amount of boxes with no values in them, then the size of the virus does not change much and is therefore weaker. The virus diffusion should be kept low for stronger viruses.

$$VD = \frac{N_v}{N_t}$$

N_v = Total number of boxes with values.

N_t = Total number of boxes.

Viral potency represents the algorithms ability to change the order of the words in the virus. How does the location of each word change after being passed through the algorithm. Does the algorithm just swap words next to each other? Or, is it capable of relocating words anywhere within the virus? Algorithms that can do this effectively will have heat maps that have low box values that are evenly distributed across the heatmap. If there is one extremely high value that is different then the surrounding values in the heatmap, then that box would stick out like a sore thumb to an antivirus software. Virial potency will be calculated by using the standard equation for statistical variance multiplied by the average values in the boxes. Both the statistical variance and average value calculations will only consider values of boxes greater than zero. The smaller the value of virial potency, the stronger the virus.

$$VP = m_v * \frac{\sum(v-m_v)^2}{N_v}$$

v = Value of each box

N_v = Total number of boxes with values greater than or equal to 1

m_v = The mean of the box with values greater than or equal to 1

I will be multiplying these two values to produce the strength of virus score. The lower the value of the strength of the virus, the better the virus.

$$VS = \frac{\sum(v-m_v)^2}{N_v} * \frac{N_v}{N_t} = \frac{\sum(v-m_v)^2}{N_t}$$

6.2 Strength of Virus Table

Virus Strength	Virus-File 3vs4	Virus-File 3vs5	Virus-File 3vs6	Virus-File 3vs7
Simple Virus	2.552971188	4.266964286	3.105180534	2.714332707
Integration Virus	7.867087576	2.187806363	2.275613276	3.433982684
GCI Virus	0.04737061937	0.01051575791	0.03783643892	0.1397128836
Permutation Virus	0.4913545884	0.8947420635	0.3266550523	0.337398374

6.3 Calculating Strength of Metamorphism

The virus strength represents how strong the virus is based on how much the virus changed between two generations. However, in order to determine how strong the metamorphic algorithm is, the variance between all of the generations of virus strength values must be assessed. This will give insight to the range of changes the metamorphic algorithm is capable of between more than just two generations of viruses. This can be done by finding the statistical variance between the different virus strengths. Before I make this calculation, I will first find the percentage change between each value of virus strength for each type of metamorphic transformation. To do this, I will divide every value of virus strength by the virus strength of virus-file 3vs4 for each technique. This just makes it so that the larger values indicate better metamorphism. From there I will calculate the statistical variance for each metamorphic technique which produces the strength of metamorphism. The larger the variance, the better the metamorphic technique.

6.4 Strength of Metamorphism Table

	Simple Virus	Integration Virus	GCI Virus	Permutation Virus
Metamorphism Strength	0.09182630073	0.1158941136	1.403513568	0.2923630428

7. Analysis

7.1 Simple Virus and Integration Virus

The Simple viruses and Integration viruses have similar characteristics and they scored 0.09 and 0.12 respectively. The enlarged heatmaps they produced had a diagonal with large values in the boxes. Furthermore, when assessing the strength of metamorphism, these two scored less than

21% of one another. The single difference between the two is that in the integration virus, the cluster of points was located in random locations while in the simple virus, the cluster was always located in roughly the same position. Unlike the other metamorphic techniques, the integration algorithm never changed the virus string which is why this virus scored so low. This score makes sense when considering how this technique is implemented in real metamorphic viruses such as the Zmist virus. This virus implemented code integration but it also implemented other techniques on top of that. Based on the strength of metamorphism scores and the heatmaps, the reason why is revealed: **solely changing the position of the virus code is not an effective method to create unique generations of malware and therefore this metamorphic technique must be supplemented with others in order to increase its strength of metamorphism.**

7.2 Permutation Virus

The permutation virus scored 1.40 which is much higher than the simple and integration viruses. Furthermore, it created heatmaps that were, by far, the most unique in comparison to the other viruses with low box values that were evenly distributed across the heatmap. The reason why it scored far better largely has to do with the permutation algorithm's ability to place words in random locations within the virus string. Each generation is dissimilar because the two virus generations will likely have their respective words placed in different random locations. While the algorithm does randomize the location of each word, there are limitations to this metamorphism. This is because the possible locations of the virus are determined by the length of the virus string. Since the virus before being transformed is of constant size, the transformed virus will also all be of constant size. Furthermore, the places where the same words appear are contained within a certain rectangular area. This rectangular area whose length and widths is the

size of the virus string limits the number of different locations the points can occupy. This limitation happens to be something that all permutation viruses have and is the reason why viruses like W32/Ghost can only create $n!$ unique generations where n represents the number of subroutines in the virus (the equivalent to the number of words in my virus string). **The permutation technique's ability to randomize the words in the virus string creates unique virus generations however, the limited degree of variance is the primary reason why this virus scores higher than the integration virus but lower than the GCI virus.**

7.3 GCI Virus

The GCI virus technique scored the most of all the viruses with a metamorphism strength of 1.40. This is because the GCI virus is particularly effective at spacing out and varying the magnitude of each space between the instances of equivalent words from each virus-file. This impacts the heatmap by increasing the number of empty spaces between each value greater than 0. The only limitation of the amount of space between each equal word is limited by the amount of garbage code that is inserted which for this investigation, was limited by the number of words in the virus-file (139 words). This means that there could be between 0 to 139 instances of garbage code insertion of different lengths that could substantially increase or decrease the distribution of the virus code across the file. These conditions are what make the GCI technique so transformative. This technique is so effective that many viruses can solely rely on an implementation of the GCI technique and still be very elusive. **So, it is because of the GCI's wide variability of virus distribution that it scores far higher than all other metamorphic techniques.**

8. Conclusion

Ever since the creation of computers malware technology has been steadily increasing in complexity and destructiveness. These developments of malware would eventually lead to the creation of metamorphic software and with it, metamorphic viruses which are a very dangerous and real threat to privacy as we know it. Though they may take years of work to develop, when they finally do get created, they have the capability of seamlessly infecting millions of computers which implement antivirus software that are based solely on static, signature based detection. Infact, if there does exist a 2021 metamorphic virus, it likely has already infected your computer without you even knowing it. This virus may be the reason why it takes just a second or two longer for a google chrome to load, or maybe could be the reason why your computer dies faster than what you expect. In any case, the only way to be sure that your computer does not have a stealthy metamorphic virus is by studying how this virus is able to transform and what methods it uses to evade detection. Studying these methods is the key to developing antivirus software that can detect metamorphic viruses.

9. Bibliography

Alam, Shahid. *A Framework for Metamorphic Malware Analysis and Real-Time Detection*.

University of Victoria, 2014,

https://dspace.library.uvic.ca/bitstream/handle/1828/5576/Alam_Shahid_PhD_2014.pdf?sequence=1&isAllowed=y. Accessed 14 Feb. 2021.

Baysa, Donabelle, "Structural Entropy and Metamorphic Malware" (2012). Master's Projects.

283. DOI: <https://doi.org/10.31979/etd.6zrt-mb7y>

https://scholarworks.sjsu.edu/etd_projects/283. Accessed 12 Feb. 2021.

Beaucampshilippe Beaucamps. *Advanced Metamorphic Techniques in Computer Viruses*.

International Conference on Computer, Electrical, and Systems Science, and Engineering

- CESSE'07, Nov 2007, Venice, Italy. Ffinria-00338066f. Accessed 18 Feb. 2021.

Duncan, Brad. "MyDoom Still Active in 2019." *paloalto*, 26 July 2019,

<https://unit42.paloaltonetworks.com/mydoom-still-active-in-2019/>. Accessed 2 Mar. 2021.

Konstantinou, Evgenios. *Metamorphic Virus: Analysis and Detection*. 15 Jan. 2008,

<https://www.ma.rhul.ac.uk/static/techrep/2008/RHUL-MA-2008-02.pdf>. Accessed 10 Feb. 2021.

Leder, Felix, Bastian Steinbock and Peter Martini. *Classification and Detection of Metamorphic Malware Using Value Set Analysis*. University of Bonn, 2009,

<https://www.computer.org/csdl/pds/api/csdl/proceedings/download-article/12OmNApcups/pdf>. Accessed 17 Feb. 2021.

- Lee, Jared, "COMPRESSION-BASED ANALYSIS OF METAMORPHIC MALWARE" (2013). Master's Projects. 329. DOI: <https://doi.org/10.31979/etd.qm6p-jkf5>
https://scholarworks.sjsu.edu/etd_projects/329. Accessed 13 Feb. 2021.
- “Oessays - The Biggest Essays Database of 2021.”. *oessays*, <https://oessays.com/>. Accessed 16 Feb. 2021. Accessed 16 Feb. 2021.
- Mishra, Puneet. *Taxonomy of Uniqueness Transformations*. San Jose State University, Dec. 2003. Accessed 16 Feb. 2021.
- Szor, Peter and Peter Ferrie. *Hunting For Metamorphic*. Symantec,
<https://crypto.stanford.edu/cs155old/cs155-spring09/papers/viruses.pdf>. Accessed 15 Feb. 2021. Accessed 15 Feb. 2021.
- Toderici, Annie Hii, "Chi-Squared Distance and Metamorphic Virus Detection" (2012). Master's Theses. 4177. DOI: <https://doi.org/10.31979/etd.j3nz-gjtr>
https://scholarworks.sjsu.edu/etd_theses/4177. Accessed 12 Feb. 2021.
- Wong, Wing and Mark Stamp. *Hunting for Metamorphic Engines*. Springer, 11 Nov. 2006,
<https://ptolemy.berkeley.edu/projects/truststc/pubs/237/hunting.pdf>. Accessed 16 Feb. 2021.
- Zhang, Qinghua. *Polymorphic and Metamorphic Malware Detection*. North Carolina State University, 2008,
<https://repository.lib.ncsu.edu/bitstream/handle/1840.16/5484/etd.pdf?sequence=1>. Accessed 10 Feb. 2021.

10. Appendix

10.1. Java Code

10.1.1. Garbage Code Insertion Algorithm

```
import java.util.Random;
import java.util.Arrays;
import java.util.List;
import java.util.ArrayList;
import java.io.File;
import java.util.Scanner;
import java.io.PrintWriter;
public class GCI
{
    private String virus = "";
    private String encVirus = "";
    public GCI(String path1, String path2)
    {
        TextVirusExtractor tve = new TextVirusExtractor(path1,
path2);
        virus = tve.getVirus();
        String[] words = virus.split(" ");
        String[] garbageTxt = {"Garbage", "Trash", "Litter",
"rubbish", "refuse", "waste"};
        ArrayList<String> wordList1 = new ArrayList<String>();
        ArrayList<String> wordList2 = new ArrayList<String>();
        for(String s : words)
        {
            wordList1.add(s);
        }
        Random ran = new Random();
        int ranInt = ran.nextInt(words.length);
        int amt;
        int ranIndex = 0;
        int totalAdded = 0;
```

```

String temp = "";
String gtxt = "";
for(int x = 0; x<ranInt; x++)
{
    amt = ran.nextInt(garbageTxt.length+1);
    ranIndex = ran.nextInt(wordList1.size());
    for(int y = 0; y<amt; y++)
    {
        gtxt+= garbageTxt[ran.nextInt(garbageTxt.length)];
    }
    for(int z = 0; z<wordList1.size(); z++)
    {
        wordList2.add(wordList1.get(z));
        if(z==ranIndex)
        {
            wordList2.add(gtxt);
        }
    }
    wordList1 = (ArrayList<String>) wordList2.clone();
    wordList2.clear();
    gtxt = "";
}
for(String s: wordList1)
{
    encVirus += s + " ";
}
encVirus = encVirus.replaceAll(" ", " ");
System.out.println(encVirus);
}

public String getVirus()
{
    File file = new File("Text1.txt");
    PrintWriter write = null;
    try
    {
        write = new PrintWriter(file);
    }
    catch(Exception ex)

```

```

    {
        System.out.println("File not found");
        System.exit(1);
    }

    write.print(encVirus);
    write.close();
    return encVirus;
}
}

```

10.1.2. Permutation Algorithm

```

import java.util.Random;
import java.io.File;
import java.util.Scanner;
import java.io.PrintWriter;
public class PermutationSwap
{
    private String virus = "";
    private String encVirus = "";
    public PermutationSwap(String path1, String path2)
    {
        TextVirusExtractor tve = new TextVirusExtractor(path1,
path2);
        virus = tve.getVirus();
        String[] words = virus.split(" ");
        String temp;
        int num1;
        int num2;
        Random ran = new Random();
        for(int x = 0; x<10000; x++)
        {
            num1 = ran.nextInt(words.length);
            num2 = ran.nextInt(words.length);
            temp = words[num1];
            words[num1] = words[num2];
            words[num2] = temp;
        }
    }
}

```

```

    }
    for(String s : words)
    {
        encVirus += s + " ";
    }
    encVirus = encVirus.replaceAll(" ", " ");
}

public String getVirus()
{
    File file = new File("Text1.txt");
    PrintWriter write = null;
    try
    {
        write = new PrintWriter(file);
    }
    catch(Exception ex)
    {
        System.out.println("File not found");
        System.exit(1);
    }

    write.print(encVirus);
    write.close();
    return encVirus;
}
}

```

10.1.3. Integration Algorithm

```

import java.util.Random;
public class NECRL
{
    private String virusText = "";
    private String text = "";
    private String virus = "";
    private int num;
    public NECRL(String path1, String path2)

```

```

    {
        TextVirusExtractor tve = new TextVirusExtractor(path1,
path2);
        text = tve.getText();
        virus = tve.getVirus();
        Random ran = new Random();
        String[] sentences = text.split("\\.");
        num = ran.nextInt(sentences.length);
        if(num<sentences.length)
        {
            for(int x = 0; x<sentences.length; x++)
            {
                if(x == num)
                {
                    virusText += virus;
                }
                virusText += sentences[x] + ".";
            }
        }
        else
        {
            System.out.print("Text is not big enough.");
            virusText = null;
        }
    }

    public String getVirusText()
    {
        System.out.println(virusText);
        return virusText;
    }
}

```

10.1.4. Compare Algorithm

```

import java.util.ArrayList;
import java.io.File;
import java.util.Scanner;

```

```

import java.io.PrintWriter;
public class SubSequenceCompare
{
    private String txt1;
    private String txt2;
    private ArrayList<Integer> m = new ArrayList<Integer>();
    private ArrayList<Integer> n = new ArrayList<Integer>();
    private ArrayList<String> z = new ArrayList<String>();
    private boolean first = true;
    private int pLength1 = 0;
    private int pLength2 = 0;
    public SubSequenceCompare(String a, String b)
    {
        txt1 = a;
        txt2 = b;
    }

    public void arrayMaker(int amt, int index1, int index2)
    {
        String sentence1 = "";
        String sentence2 = "";
        int Lindex1 = index1;
        int Lindex2 = index2;
        int temp1 = 0;
        int temp2 = 0;
        if(first)
        {
            index1 = 0;
            index2 = 0;
            first = false;
        }
        else
        {
            Lindex1++;
            Lindex2++;
        }
        while(temp1 <= amt || temp2 <= amt)
        {

```

```

        if(index1<txt1.length() && txt1.substring(index1,
index1+1).compareTo(".")==0)
        {
            sentence1 = txt1.substring(Lindex1 ,index1);
            temp1++;
            index1++;
        }
        else if (index1>txt1.length())
        {
            temp1++;
        }
        if(temp1 <= amt)
        {
            index1++;
        }

        if(index2<txt2.length() && txt2.substring(index2,
index2+1).compareTo(".")==0 )
        {
            sentence2 = txt2.substring(Lindex2 ,index2);
            temp2++;
            index2++;
        }
        else if(index2>txt2.length())
        {
            temp2++;
        }
        if(temp2 <= amt)
        {
            index2++;
        }
    }
    //System.out.println(sentence1);
    this.comparer(sentence1, sentence2);
    pLength1+= sentence1.length()+2;
    pLength2+= sentence2.length()+2;
    //System.out.println(pLength2);
    //System.out.println(pLength1);
    if(index1>txt1.length()-1 || index2>txt2.length()-1)

```

```

        {
            return ;
        }
        else
        {
            this.arrayMaker(amt, index1, index2);
        }
    }

    private void comparer(String sentence1, String sentence2)
    {
        String word1 = "";
        String word2 = "";
        int start1 = 0;
        int start2 = 0;
        boolean wordfound = false;
        for(int x = 0; x<sentence1.length()-1; x++)
        {
            if(sentence1.substring(x, x+1).compareTo(" ")==0 ||
            ((sentence1.substring(x, x+1).compareTo(".")==0) &&
            (sentence1.substring(x+1, x+2).compareTo(" ")==0)))
            {
                word1 = sentence1.substring(start1, x);
                wordfound = true;
                x++;
                //System.out.println(start1);
            }

            for(int y = 0; y<sentence2.length()-1 && wordfound; y++)
            {
                if(sentence2.substring(y, y+1).compareTo(" ")==0 ||
                ((sentence2.substring(y, y+1).compareTo(".")==0) &&
                (sentence2.substring(y+1, y+2).compareTo(" ")==0)))
                {
                    word2 = sentence2.substring(start2, y);
                    // System.out.println("2" + pLength2);
                    //System.out.println("1" + pLength1);
                    y++;
                    if(word1.equalsIgnoreCase(word2))

```



```

        {
            m.add(start1+pLength1);
            n.add(start2+pLength2);
            z.add(word1);
            start2 = y;
        }
        else
        {
            start2 = y;
        }
    }

    word2 = "";
    if(sentence2.substring(y, y+1).compareTo(" ")==0)
    {
        y++;
        start2++;
    }
}
if(wordfound)
{
    start1 = x;
}
wordfound = false;
start2 = 0;
if(sentence1.substring(x, x+1).compareTo(" ")==0)
{
    x++;
    start1++;
}
}

}

public void returnArray()
{
    String str = "";
    int num = 0;
    while(num<m.size())
    {

```

```

        str = num+1 + ": (" + m.get(num) + ", " + n.get(num) + ")
" + z.get(num);
        System.out.println(str);
        num++;
    }
}

public void returnArrayData(String path1, String path2)
{
    File file1 = new File(path1);
    File file2 = new File(path2);
    String str = "";
    int num = 0;
    PrintWriter write1 = null;
    try
    {
        write1 = new PrintWriter(file1);
    }
    catch(Exception ex)
    {
        System.out.println("File not found");
        System.exit(1);
    }

    PrintWriter write2 = null;
    try
    {
        write2 = new PrintWriter(file2);
    }
    catch(Exception ex)
    {
        System.out.println("File not found");
        System.exit(1);
    }
    while(num<m.size())
    {
        str = m.get(num) + "";
        write1.println(str);
        System.out.println(str);
    }
}

```

```

        num++;
    }
    write1.close();
    System.out.println("*****");
    num = 0;
    str = "";
    while(num<n.size())
    {
        str = n.get(num) + "";
        write2.println(str);
        System.out.println(str);
        num++;
    }
    write2.close();
}
}

```

10.2. Texts

10.2.1. Virus String

I am a virus that has just infected this text. Because of me, this data will likely become corrupt and unusable. Hopefully, because of my properties, I will be able to blend in and not be detected. If I am lucky, the similarity tests that I have made will not detect me. If not, then that will not be fun and I will be very sad. If I want to, I can completely destroy this file and all its contents. I can also spread to other files that I have made however I have to do so manually. The program this virus is programmed in is object oriented java. Java is great because it allows for great organization along with many other things. Well, I am going to stop the virus text because it is now becoming too long.

10.2.2. Essay 3

Big Data has been created lot of buzz in Information Technology word. Big Data contain large amount of data from various sources like Social Media, News Articles, Blogs, Web, Sensor Data and Medical Records etc.

Big Data includes Structured, Semi-Structured and Unstructured data. All these data are very useful to extract the important information for analytics.

Big Data is differs for other data in 5 Dimensions such as volume, velocity, variety, and value. Big Data include social media, Product reviews, movie reviews, News Article, Blogs etc.. So, to analyze this kind of unstructured data is challenging task.

This thing makes Big Data a trending research area in computer Science and sentiment analysis is one of the most important part of this research area.

As we have lot of amount of data which is certainly express opinion about the Social issues, events, organization, movies and News which we are considering for sentiment analysis and predict the future trends and effect of certain event on society.

We can also modify or make the improve strategy for CRM after analysing the comments or reviews of the customer. This kind analysis is the application of Big Data.

Big Data is trending research area in computer Science and sentiment analysis is one of the most important part of this research area. Big data is considered as very large amount of data which can be found easily on web, Social media, remote sensing data and medical records etc. in form of structured, semi-structured or unstructured data and we can use these data for sentiment analysis.

Sentimental Analysis is all about to get the real voice of people towards specific product, services, organization, movies, news, events, issues and their attributes

Sentiment Analysis includes branches of computer science like Natural Language Processing, Machine Learning, Text Mining and Information Theory and Coding. By using approaches, methods, techniques and models of defined branches, we can categorized our data which is unstructured data may be in form of news articles, blogs, tweets, movie reviews, product reviews etc. into positive, negative or neutral sentiment according to the sentiment is expressed in them.

Document Level Sentiment analysis is performed for the whole document and then decide whether the document express positive or negative sentiment.

Entity or Aspect Level sentiment analysis performs finer-grained analysis. The goal of entity or aspect level sentiment analysis is to find sentiment on entities and/or aspect of those entities.

For example consider a statement “My HTC Wildfire S phone has good picture quality but it has low phone memory storage.” so sentiment on HTC’s camera and display quality is positive

but the sentiment on its phone memory storage is negative. We can generate summary of opinions about entities. Comparative statements are also part of the entity or aspect level sentiment analysis but deal with techniques of comparative sentiment analysis.

Sentence level sentiment analysis is related to find sentiment from sentences whether each sentence expressed a positive, negative or neutral sentiment. Sentence level sentiment analysis is closely related to subjectivity classification. Many of the statements about entities are factual in nature and yet they still carry sentiment. Current sentiment analysis approaches express the sentiment of subjective statements and neglect such objective statements that carry sentiment. For Example, “I bought a Motorola phone two weeks ago. Everything was good initially. The voice was clear and the battery life was long, although it is a bit bulky. Then, it stopped working yesterday.

” The first sentence expresses no opinion as it simply states a fact. All other sentences express either explicit or implicit sentiments. The last sentence “Then, it stopped working yesterday” is objective sentences but current techniques can’t express sentiment for the above specified sentence even though it carry negative sentiment or undesirable sentiment. So I try to solve out the above problematic situation using our approach.

The Proposed classification approach handles the subjective as well as objective sentences and generate sentiment from them.

The objective of this research work is to improve the effectiveness and efficiency of classification as well as sentiment analysis because this analysis plays a very important role in analytics application.

Till now Sentiment analysis focus on Subjectivity or Subjective sentiment i.e. explicit opinion and get idea about the people sentiment view on particular event, issue and products. Sentiment analysis does not consider objective statements although objective statements carry sentiment i.e. implicit opinion.

So here the main objective is to handle subjective sentences as well as objective sentences and give better result of sentiment analysis.

Classification of unstructured data and analysis of classified unstructured data are major objectives of me.

Practical implementation will be also done by me in the next phase.

Scope of this dissertation is described as below.

There are so many application of Sentiment Analysis which is used now-a-day to generate predictive analysis for unstructured data.

Areas of applications are Social and Business intelligence applications, Product reviews help us to define marketing or production strategies, Movie reviews analysis, News Analysis, Consider

political news and comments of people and generate the analysis of election, Predict the effect of specific events or issues on people, Emotional identification of person can be also generated, Find trends in the world Comparative view can also be described for products, movies and events, Improve predictive analysis of return of investment strategies.

10.2.3. Essay 4

SWOT is analysis of company. It is opened as Strengths, Weakness, Opportunities and Treats. With this model you can analyze what can or cannot do the company, and also what are the potential opportunities and threats. This is environmental analysis, which include external and internal issues. As the pioneer of this analysis, and one of the first strategy theorists Ken Andrew was the first who analyzed the strategy with considering capabilities and resources with the external environment. For analyzing and strategizing we should consider the following points: As Strength and Weakness we should answer the question: what can we do? What resources and capabilities do we want to develop? What do we want to do? And what should we be carrying about? (Organizational and context attributes). As external Opportunities and Threats we should answer the question: what might we do? What opportunities can we develop? Considering what do other expect us to do? How can we build shared expectations among our stockholders? First you need to diversify and classify by ranking the environmental issues, for example for internal strengths write first most important, than 2nd than so on. Second you should put together the internal factors with external factors, example showed below.

Here you can see how to match them, 1. The internal strength is analyzed with external opportunities, 2nd internal weakness relative to external opportunities and so on. With the help of this you consider all factors that have and/or could affect the company, so with this method you can formalize strategy for company.

Now we will see some methods or tools to gather strengths, weaknesses, opportunities and threats.

Industry Analysis Perspectives: this analysis includes structural analysis, competitive forces analysis and supply chains & value chains analysis.

Structural analysis focuses on external structural variables, which affects firms' strategy and performance. This analysis contains structure, conduct and performance. The last aspects can be considered as internal factor: this analysis representing and analyzing the industry with 5 forces, in the middle of industry competitors, in one side are buyers, in the opposite site are suppliers and the other two are substitutes and new entrants. This analysis can be considered as external factors.

Competitive map based on resource and market commonality: This is a great tool which helps to identify current and potential competitors and find the position of the company in the competitive situation. This tool will help to analyze the external factors. It will show what is current competitive situation in the market (treats or internal strengths) and what opportunities it has.

Cross price elasticity: This is good way to identifying competitors through interdependency. This help to identify the interactions with order markets and also identifies the players in the same market. Cross price elasticity tool can help to analyze the external factors of company and also internal strengths. Structural homogeneity: In this analysis you look at different strictures and see which of them matches with competitors. This tool we can consider as external factors, but they include the internal factors also, cause the position of company in the market is coming from the internal environment.

Product differentiation: This is measuring the product with individual characteristics, considering all characteristics of a product, not the product as one object. There are two ways of differentiation, one is vertical differentiation, and the other is horizontal differentiation. This tool can help to SWOT analysis to measure the external factors, how good is product in the market.

Importance/Performance analysis: This is analysis of company's internal factors. It is measuring the performance of the factor and importance, thus you can see the strengths or weakness of the internal organization.

The evaluation of overall satisfaction: This is tool to evaluate the satisfaction from characteristic of product with the help of Kano model questionnaire. Represented in two dimensions, one is satisfaction/dissatisfaction and the other is characteristic value. This is one of the ways to evaluate the internal factors of company.

Competitive map: This is a good tool to analyze market position of the product or company. It is competitive map based on market share and relative price. It is represented in two dimensions, one is relative price the other dimension is market share. This is a very good tool to analyze product relatively to competitors, it can support to SWOT analysis for external factors.

System mapping is representing cause-effects of the events. With the help of this you can make the strategy for company, it include both internal and external factors, so this tool can be joined to SWOT analysis to help to formalize strategy for company.

Strategic Business Unit: This is competitive analysis through homogeneity. You should take tree factors from which you make graph with tree dimensions. The tree factors are technology, benefits and group of customers. From this analysis you can conclude both internal and external factors, so this is another way, which we can put horizontally with SWOT analysis, it is we can say "substitute" but also can be support to SWOT analysis.

Price sensitivity and product differentiation: This is analyzing the product's place in the market considering the quality and price. The representation of this analysis is two dimensions, one is relative expected price the other side is perceived quality. With the help of this you can see the place in the market with quality and price as a measure. This is good tool to analyze the internal strength and/or weakness and comparison with competitors. I am considering this as an internal measure, And afterwards you can see the threat from other company's and opportunities in the market.

We have looked threw some tools and methods which can support to SWOT analysis. I have classified the, as internal and external or general analysis, because based on analysis only you can know the result is strength or weakness for your company. Below I am representing them once more together identified, the tools that are not included are general analysis, which contains both internal and external factors.

10.2.4. Essay 5

Strategic analysis is one of the most valued procedures in business. It is widely used by many of business because of its rewarding effects to the firm or organization. Strategic analysis involves the analysis of markets, consumer, costs, competition and other aspects of the business.

There are a lot of different methods and techniques when it comes to strategic analysis. Some use the PEST method, and some use the SWOT method. These two are the most common used method in strategic analysis. However, there are others such as Scenario planning, five forces analysis, market segmentation, competitor analysis, and many more.

The PEST analysis is usually utilized when a person wants to analyze the 'environment' in which the business is operating. Of course, it is important to consider first the environment before proceeding with marketing procedures (market segmentation etc.). A firm's environment is made up of three sub-environments. These are the internal environment, micro-environment, and the macro-environment.

The internal environment includes the staff and other personnel, wages, office equipment, and other financial matters. The micro-environment is made up of customers or the consumers, distributors, suppliers, agents, competitors, and other people outside the firm but has a direct impact on the organization. The examples of macro-environment are as follows: political, economic, sociocultural, and technological. This is what we call now the PEST analysis. The PEST analysis tries to analyze these factors of the macro-environment (2000).

Another effective technique in strategic analysis is market segmentation. Although many firm find it hard to use this kind of method, this is one of the most effective when done correctly.

Market segmentation enables a firm to identify the similarities and differences between group of consumers or customers. However, there has to be criteria in segmenting markets. What makes it much harder to use this technique is that the criteria used in the consumer market are different from that of the business market.

The criteria for consumer market segmentation are different from the ones used in business markets. There are four major criteria for consumer market segmentation. They are geographic, demographic, psychographic, and behavioral segmentation. Business market segmentation, on the other hand, uses five criteria namely demographic, operating variables, purchasing approaches, situational factors, and personal characteristics (Kotler, 2003).

Geographic segmentation involves the division of market into various geographical units such as states, counties, regions, cities, and municipalities. Since consumer preference differs from one area to another, geography is a great consideration. For example, in the case of Hilton hotels, room and lobby designs are based according to the location of the hotel. Northeastern hotels are sleeker and more sophisticated while Southeastern hotels use a “rugged” or rustic approach.

Aside from location, there are still a lot of factors that have to be considered. They are the size of the certain area, density, and climate. In demographic segmentation, the factors considered are age, family size, family life cycles, gender, income, occupation, education, religion, race, generation, nationality and social class.

Demographic factors are the most common criteria for segmenting markets simply because they are usually connected with consumer preferences, wants, and usage rates. Psychographic factors include lifestyle and personality of consumers. Since people have different lifestyles, it is certain that they also have different wants and preferences. Consumers differ in attitudes, interests, and activities and they may have great effects on the goods and services they consume. For example, a person that is into sports may consume more training or gym services than the person who is not. On the other hand, personality is also one of the factors that have to be considered in market segmentation. Because of this, products also tend to have their own “attitudes”. Nike, for example, wishes to appear exciting while Levi’s wanted a youthful, rugged, and rebellious attitude. Behavioral segmentation, obviously concerns about consumer behavior. The usual considerations with this criterion are occasions, benefits, user status, usage rate, and the attitude of the consumer towards the product (Kotler, 2003).

Just like in the consumer market, demographic factors are also considered by business markets when it comes to segmentation. But unlike the consumer markets, demographic factors in the business market include industries, companies, and geographical areas that they should serve. The operating variables that have to be considered in this market are the customer technologies

and the user or non-user status that the firm should focus on. Customer capabilities are also considered (whether the firm should serve customers needing many or few services).

Purchasing approaches include purchasing-function organization, power structure (engineering dominated or financially dominated etc.), nature of existing relationships with other companies, general purchasing policies (includes leasing, bidding, contracts, purchases) and lastly, purchasing criteria (involves quality, Service, price). Considerations of the situational factors include urgency, specific applications, and size of order. The last criterion for segmenting market is the personal characteristics. Factors that have to be taken into consideration are buyer-seller similarity, attitudes toward risk, and loyalty (Kotler, 2003).

Another method of strategic analysis is the SWOT analysis. This method enables organizations to assess the business and its environment by evaluating the strengths, weaknesses, opportunities, and threats (giving the name SWOT) of the business scenario. Because of its positive outcomes, this method is widely used in analyzing business cases in several organizations as well as in the academe.

The strengths of the organization include skills, current market position, patents, and the other competencies of the firm or organization. On the other hand, the weaknesses are the factors that can put the company into a bad light. This could be poor performance or anything that can do harm to the organization. This includes poor management, poor technological equipment, debts, and others. Opportunities are factors (current or future) that can be used by the company to their advantage, Threats, however, can be current or future conditions which can harm the organization. Although the SWOT analysis does not give exact answers, it is one of the widely used because of its usefulness when it comes to organizing information and somehow predicting good and bad events. In this way, SWOT analysis is a great help in making business plans and operations (2007).

The three methods of strategic analysis explained above are only a few of the methods that are really helpful in working towards effective strategic management. We already mentioned above the contributions of the specific methods to effective strategic management. Now we have to site the contributions of strategic analysis in general to effective strategic management. The figure below shows the three components of strategic management process.

With this process, effective strategic management is attained. Strategic analysis helps the organization in identifying its environment and analyzing the factors that affects that certain business environment. This was provided to us by the PEST analysis. It also gives the organization a more realistic goals and objectives that go with the missions of the company. It also gives strategies that go well with the organization's capacity to implement.

Another benefit that we can get from strategic analysis which can lead to effective strategic management it makes more efficient and effective use of the company's existing resources in carrying out the organization's objectives. The use of strategic planning can make more room for progress since it gives the information that a firm needs especially when it involves change. Lastly, in using strategic analysis and planning, the organization build a clearer direction for the company which results to effective strategic management (1997).

10.2.5. Essay 6

A job analysis is a way to describe a job or the attributes needed to obtain the job (Spector, 2012). Different jobs need people with certain attributes; these attributes help the job be completed as necessary. A job analysis can help companies hire the right employees, and this creates a successful working environment. There are two different job analysis approaches; each approach focuses on different needs for a job. The first is the Job-Oriented approach; this approach describes the nature of the job, some job analysis, go into specific task (Spector, 2012). The second is the Person-oriented approach; this approach describes the specific attributes or characteristics needed to successfully do the job (Spector, 2012). This paper will conduct a job analysis for my preferred job using one of the job analysis methods and discuss how it could be used within an organization. Evaluate the reliability and validity of the job analysis. Evaluate different performance appraisal methods that might be applied to my chosen job and explain the various benefits and vulnerabilities of each performance appraisal method.

Job Analysis: School Psychologist To analyze the job that I would dream to have I will be using both the Job-oriented approach and the person-oriented approach. I will be describing both the task and attributes a school psychologist needs. To start off with the job-oriented approach, which describes the jobs nature and specific tasks, I will describe what being a school psychologist is all about. School psychologists help children and youth succeed academically, socially, behaviorally, and emotionally.

They collaborate with educators, parents, and other professionals to create safe, healthy, and supportive learning environments that strengthen connections between home, school, and the community for all students. School psychologists not only work with students but they also provide services for families, teachers, administrators, and the community. **Job-Oriented Approach** According to the National Associations of School Psychologist the following are tasks that school psychologist need to complete to be successful job.

When it comes to students they Provide counseling, instruction, and mentoring for those struggling with social, emotional, and behavioral problems. Increase achievement by assessing

barriers to learning and determining the best instructional strategies to improve learning. Promote wellness and resilience by reinforcing communication and social skills, problem solving, anger management, self-regulation, self-determination, and optimism. Enhance understanding and acceptance of diverse cultures and backgrounds.

They also help out families by identifying and addressing learning and behavior problems that interfere with school success. Evaluate eligibility for special education services, support students' social, emotional, and behavioral health, also teach parenting skills and enhance home-school collaboration, and make referrals and help coordinate community support services. School psychologist also provides a support for teachers by identifying and resolving academic barriers to learning. Design and implement student progress monitoring systems and academic and behavioral interventions.

Support effective individualized instruction, create positive classroom environments, and motivate all students to engage in learning. School psychologist also give support to administrators by collecting and analyzing data related to school improvement, student outcomes, and accountability requirements. Implement school-wide prevention programs that help maintain positive school climates conducive to learning. Promote school policies and practices that ensure the safety of all students by reducing school violence, bullying, and harassment.

Respond to crises by providing leadership, direct services, and coordination with needed community services. There are many tasks that come with being a school psychologist; they provide support for many different departments of a school, and take on tasks that are delicate. Person-Oriented Approach Person-oriented approach lists the characteristics and attributes needed to be a school psychologist. Since this job is very serious and deals with people's mental and emotional health, the first and most important attribute is education.

The job needs extensive knowledge of the field and practice before being able to qualify for this job. Other requirements are positive attitude, loving the job, loving to help people, and love psychology. They also have to be focused and responsible because if they aren't an accident can occur. The person should also be able to keep their personal life and their job separate. If they cannot do this their job can affect their personal job or vice versa. From the job tasks these are the attributes that an individual has to have to be able to be successful at this job.

There are many job analysis types and not all are appropriate for the type of job application (Spector, 2012). Each has its limitations and advantages, and it works for different purposes. The reliability and validity of a job analysis can be questioned because the analysis is made by a person who is either doing the job or watching someone do it (Spector, 2012). The biggest thing that we do as human beings is judge, and judgment can make a job analysis not reliable or valid.

In this case I am not a school psychologist, and I am actually reading about the job, my judgment can have a big impact on the job analysis.

Validity and reliability are questionable matter but as quoted “In general, results suggest that different people’s ratings of jobs are often reasonably reliable. In other words, there will be a relatively high correlation among different people’s ratings of the same job for at least some job analysis methods (Spector, 2013 p. 72-73). ” In other words some job analysis methods may have the same ratings for certain jobs. Appraisals in the work place can be very time consuming for supervisors but it has a good outcome, appraisals create a better performance in the work place.

When someone is appraised for their good work they get satisfaction, and that’s something they want to keep receiving so they keep up the good work. For my chosen job I think that appraisals should be given, because a school psychologist makes people feel good and therefore they should get appraisal to feel good about what they are doing. Conclusion Becoming a school psychologist you have to have a lot of qualifications and characteristics that will make you successful at this job, and making sure that you know what it takes and if you are qualified for the job can be determined by a job analysis.

There are many methods to conduct a job analysis, you can use the job-oriented approach or the person-oriented approach, only two of the many. These types of analysis can explain if you are right for a job. This paper conducted a job analysis for a school psychologist job by using both job-oriented and person-oriented approach and discussed how it could be used within an organization. Evaluated the reliability and validity of the job analysis. Evaluated different performance appraisal methods that might be applied to my school psychologist job and explain the various benefits and vulnerabilities of each performance appraisal method.

10.2.6. Essay 7

Situational analysis is defined as a methodical compilation and assessment of past and current economical, political, social, and technological data. This has an objective of assessing the organizations PEST and SWOT analysis, and also identification of internal and external forces that may affect the company’s performance and choosing of strategies. Market opportunity analysis is the utilization of forecasting techniques to the market factors that may influence the demand a product.

It is also defined as formal ordered effort to recognize future situations which can be utilized to gain competitive advantage. Several techniques ranging from simple extrapolation complex technological forecasting and decision support systems can be used in market analysis. The

purpose of market opportunity analysis is to determine if there is sufficient demand for the envisioned product or service and to establish if the product or service can be offered on a profitable basis in the market as argued by Boar (2010).

The function of market opportunity analysis is to synchronize the operations of the company with the goals and objectives in order to gain competitive advantage. Task Information Systems is an organization with twenty three years experience whose business is the provision of strategic information systems for organizations, the company offers several Information and Communication technology based services such as M. I. S. / I. T. Department Management and Technical Project Management, Information Technology consultancy, network design, tailor made software, web development and user Training.

A customer market opportunity analysis at Task information systems involves analysis on the firm's priorities if they are in agreement with the organizations goals and objectives. The company intends to be a world class player in the information systems market, however, its objective is to provide information systems to fifty companies in two years. An analysis reveals that the goal and the objectives may not be in agreement since to be a world class player requires offering services to more customers in a year, not just fifty in two years.

The company needs to clearly segregate customers, so as to be understandable on the criterion of determining the fifty companies to be served in two years. An analysis on the flexibility of the company in terms of willingness to tailor services to suit the needs of the customers, involves working with customers to design and redesign products. An analysis of the firm's strengths reveals that the company has good community customers, moreover Task information Systems is a relatively flexible organization which can be attested by its strategy on customer satisfaction. Another important concept on market opportunity analysis is on the products offered by the company, if they can be supported over a long duration of time for a number of targeted customers. The company offers M. I. S. / I. T. Department Management and Technical Project Management, Information Technology consultancy, network design, tailor made software, web development and user Training among several other services, this can be compared with the available knowledge base and expertise.

The analysis reveals that the company is promising more than it can deliver to the satisfaction of customers, since if it is to offer all the advertised services the workforce would be overwhelmed. An analysis on ways in which the organization can make better, more productive use of the information to which is already in its domain to create effective customer satisfaction and relations, reveals that the company has the opportunity of faster information, that allows business with no boundary limit in terms of potential customers, furthermore the organization has positioned itself to deal with the latest innovations in the market.

Task information utilizes the latest innovations such as Search Engine Positioning, NetObjects Fusion, Sonicwall, ZoneAlarm, and Kerio Firewalls this therefore reveals that the company is dynamic and stands to benefit from its strong capacity to consolidate on various innovations. On analysis of the existing policies, procedures and practices governing the management and utilization of information throughout the organization, reveals that information utilization in the organization is not centralized and there exists no centralized management structure, this reveals therefore that the company lacks good managers with good leadership skills, which may affect customer experience and relations.

An analysis on the opportunities available for Task information systems, which can be utilized to gain competitive advantage include deregulation allows the company to advance into new markets hence more customers, in addition, the company can utilize the opportunity of faster information allows business with no boundary limit, this is because most of the services, offered by the firm are internet based, therefore the company can reach several clients at a time.

Another present opportunity is in the strategies utilized by the organization to gain a market niche, the company utilizes added value, customer satisfaction, training, aggressive marketing, e-commerce and lower cost, in order to increase the customer base and gain competitive advantage. Conclusion To gain competitive advantage therefore, Task information systems has to focus on effectively identifying and meeting the unmet/underserved customer needs, and eliminate barriers that block some/all potential customers, while at the same time innovating opportunities that enhance the customers' experience.

The strategic moves to be taken are value chain which is a set of value creating activities within the company. This includes trapped value which is concerned with efficiency, accessibility and customer empowerment. New value involves personalization, extension, community-building, collaboration while horizontal plays improve functional operations to satisfy customer experience, in addition the company should adopt vertical plays which in concerned with improving industry specific business activities. Importance of gap analysis, determining weakness in other organizations.

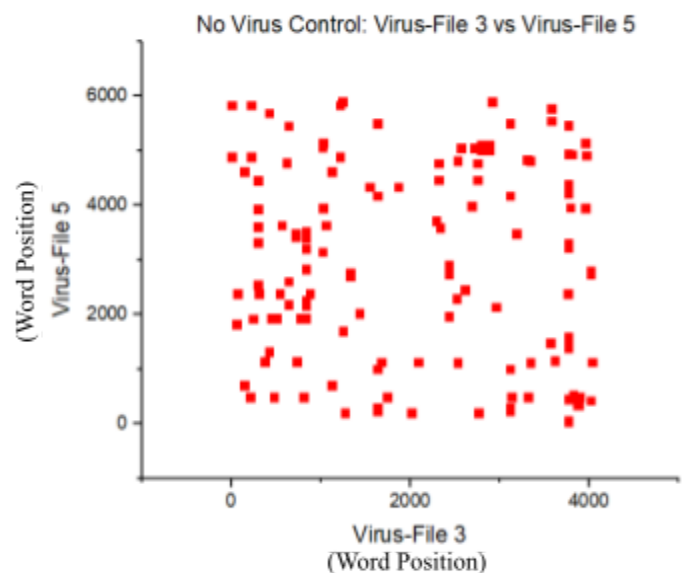
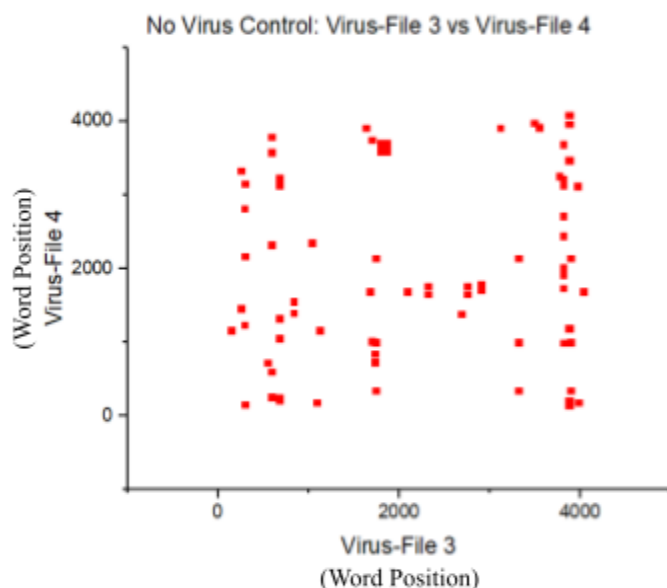
Determining weakness in other organization is very important to assist the organization in deciding upon marketing strategies and tactics, knowing the weakness of the competitors is beneficial in understanding how to exploit those weaknesses while marketing the products of the company and pointing out the strengths of the company to the customers. Determining weaknesses of competitors is also important to assist the organization to avoid those weaknesses, and avoid actions and situations that may lead to the same weaknesses.

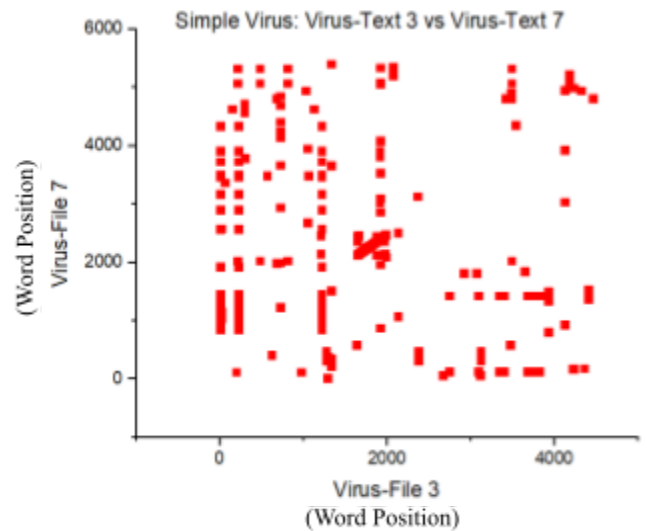
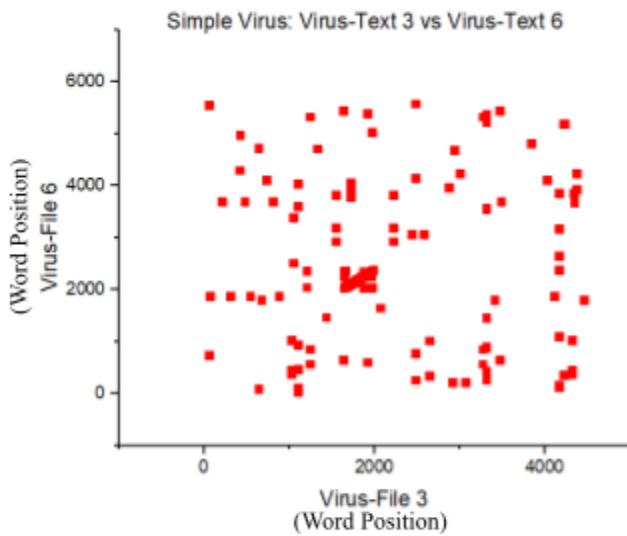
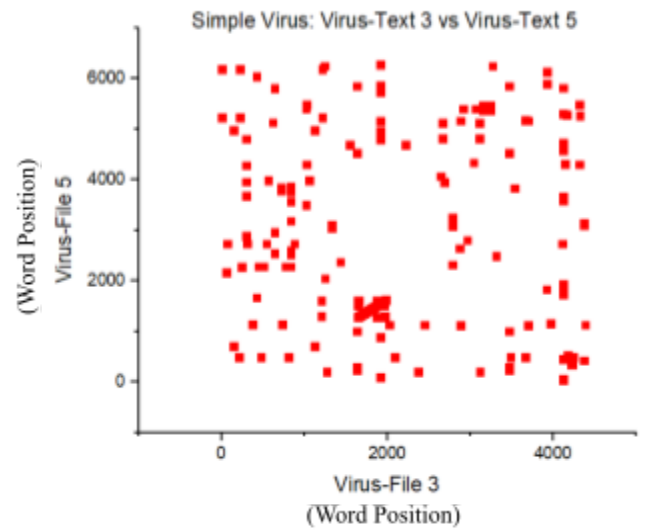
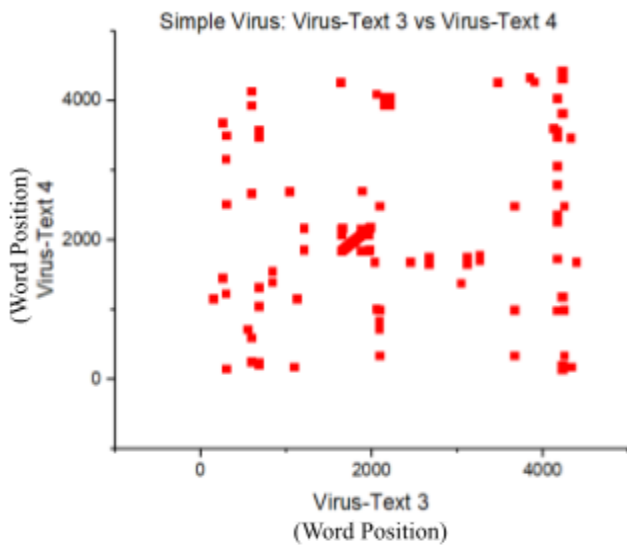
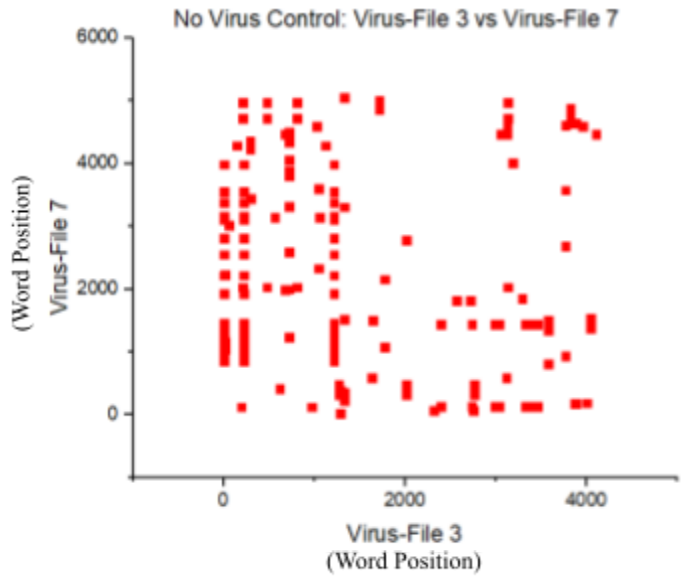
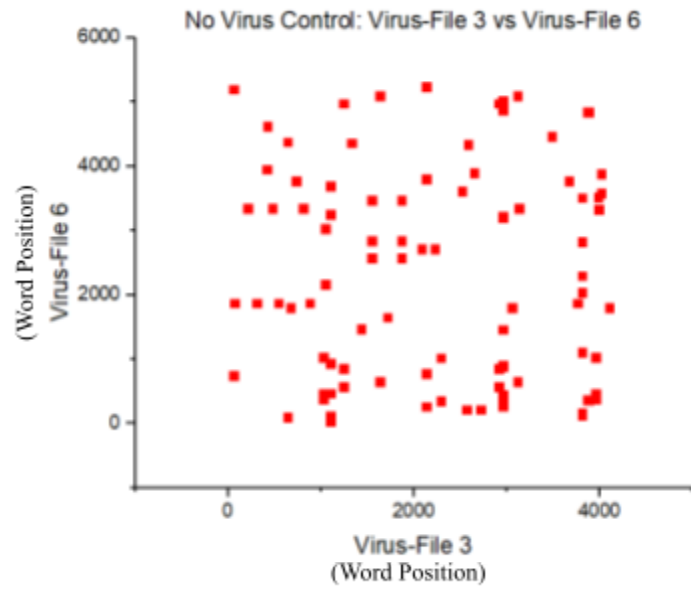
big created buzz information technology word. contain large amount various sources social media, news articles, blogs, web, sensor medical records etc. big includes structured, semi-structured unstructured data. extract important information analytics. big differs 5 dimensions volume, velocity, variety, value. include social media, reviews, movie reviews, news article, blogs etc.. so, kind unstructured challenging task. this thing makes trending research area computer science sentiment important part research area. amount certainly express opinion social issues, events, organization, movies news considering sentiment predict future trends effect certain event society. we modify improve strategy crm analysing comments reviews customer. kind application data. big trending research area computer science sentiment important part research area. considered large amount found easily web, social media, remote sensing medical records etc. form structured, semi-structured unstructured data sentiment analysis. sentimental real voice towards specific product, services, organization, movies, news, events, issues attributes sentiment includes branches computer science natural language processing, machine learning, text mining information theory coding. using approaches, methods, techniques models defined branches, categorized unstructured may form news articles, blogs, tweets, movie reviews, reviews etc. positive, negative neutral sentiment according sentiment expressed them. document level sentiment performed whole document decide whether document express positive negative sentiment. entity aspect level sentiment performs finer-grained analysis. virus infected text. me, likely become corrupt unusable. hopefully, properties, able blend detected. lucky, similarity tests detect me. not, fun sad. to, completely destroy file contents. spread files however manually. program virus programmed object oriented java. java allows organization along things. well, going stop virus text becoming too long. goal entity aspect level sentiment find sentiment entities and/or aspect those entities. for example consider statement “my htc wildfire s phone picture quality low phone memory storage.” sentiment htc’s camera display quality positive sentiment phone memory storage negative. generate summery opinions entities. comparative statements part entity aspect level sentiment deal techniques comparative sentiment analysis. sentence level sentiment related find sentiment form sentences whether each sentence expressed positive, negative neutral sentiment. sentence level sentiment closely related subjectivity classification. statements entities factual nature carry sentiment. current sentiment approaches express sentiment subjective statements neglect objective statements carry sentiment. for example, “i bought motorola phone weeks ago. everything initially. voice clear battery life long, although bit bulky. then, stopped working yesterday. ” sentence expresses opinion simply states fact. sentences express either explicit implicit sentiments. last sentence “then, stopped working yesterday” objective sentences current techniques can’t express sentiment above specified sentence though carry negative sentiment undesirable sentiment. try

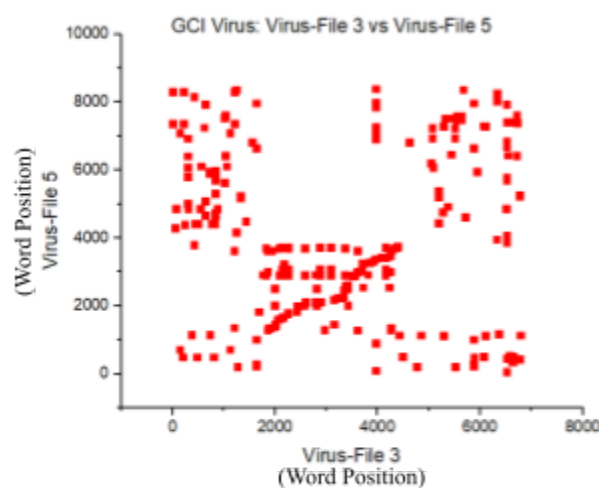
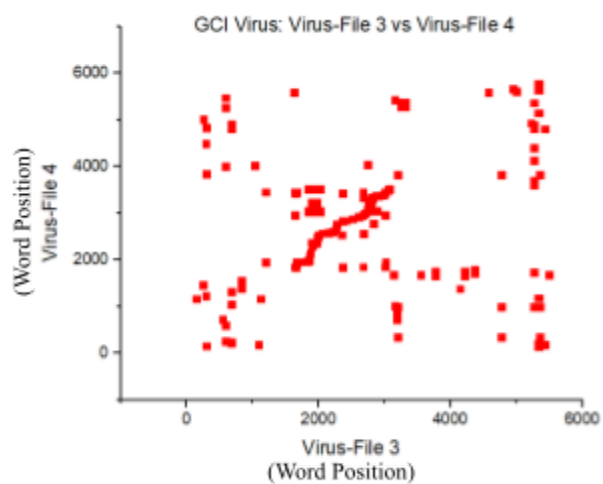
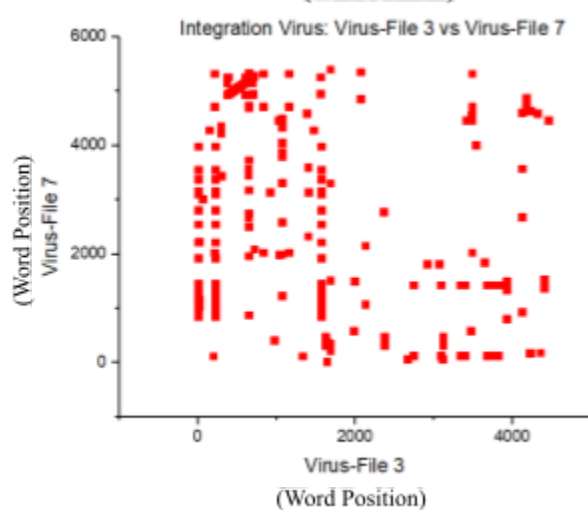
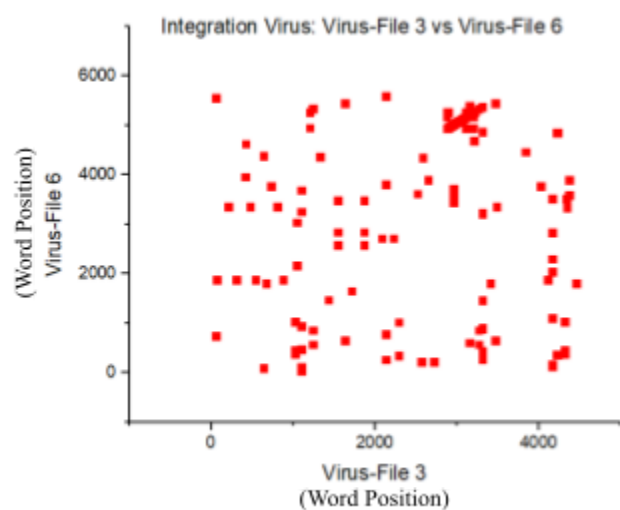
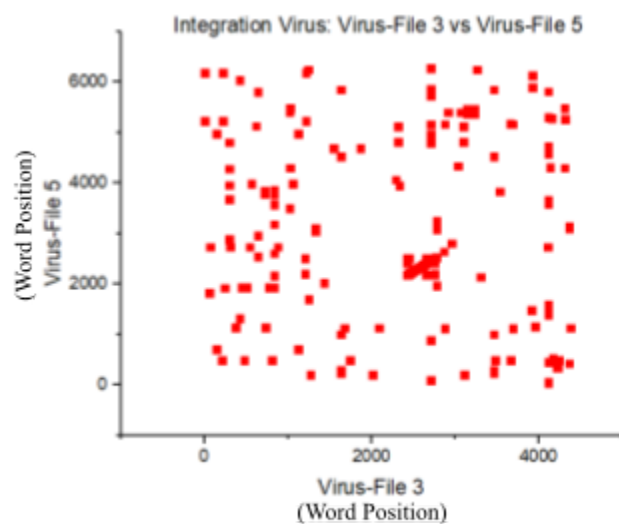
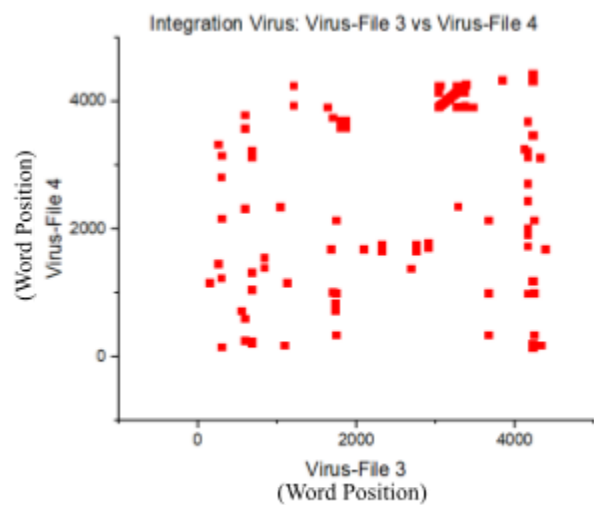
solve above problematic situation using approach.the proposed classification approach handles subjective objective sentences generate sentiment form them.the objective research improve effectiveness efficiency classification sentiment analysis plays important role analytics application.till sentiment focus subjectivity subjective sentiment i.e. explicit opinion idea sentiment view particular event, issue products. sentiment does consider objective statements although objective statements carry sentiment i.e. implicit opinion.so here main objective handle subjective sentences objective sentences better result sentiment analysis.classification unstructured classified unstructured major objectives me.practical implementation done next phase.scope dissertation described below.there application sentiment now-a-day generate predictive unstructured data.areas applications social business intelligence applications, reviews help define marketing production strategies, movie reviews analysis, news analysis, consider political news comments generate election, predict effect specific events issues people, emotional identification person generated, find trends world comparative view described products, movies events, improve predictive return investment strategies.

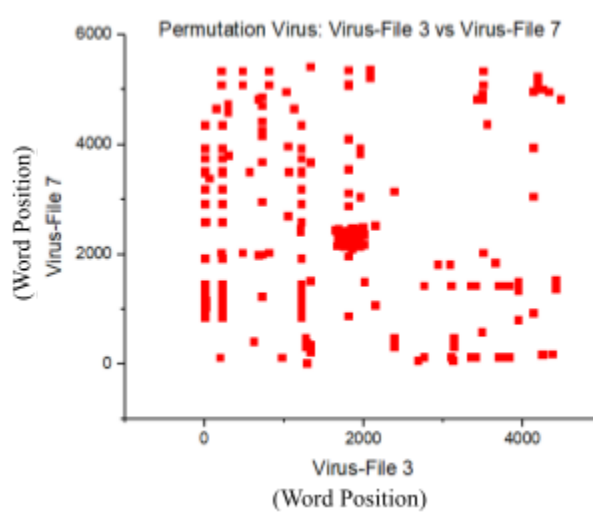
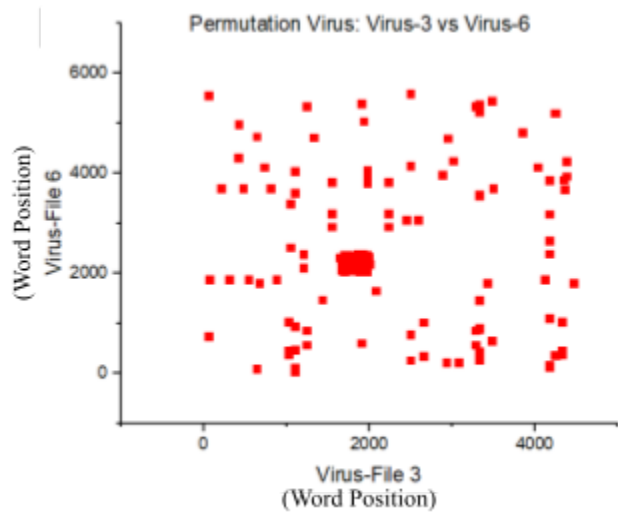
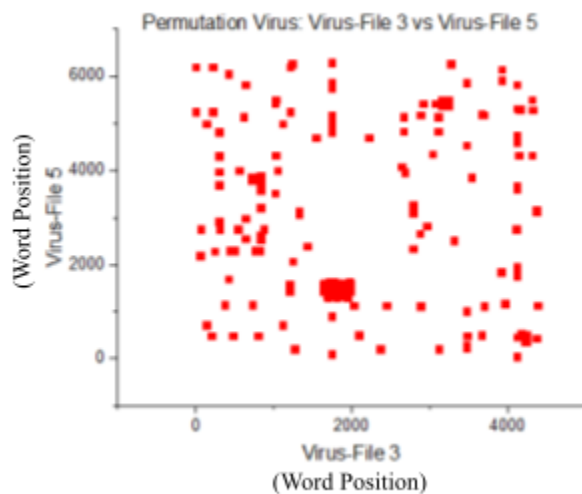
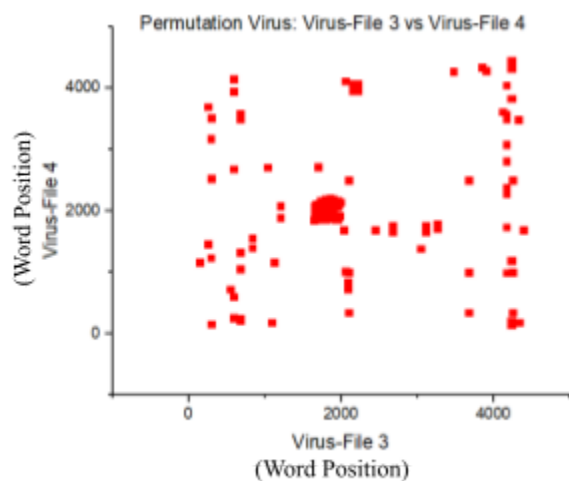
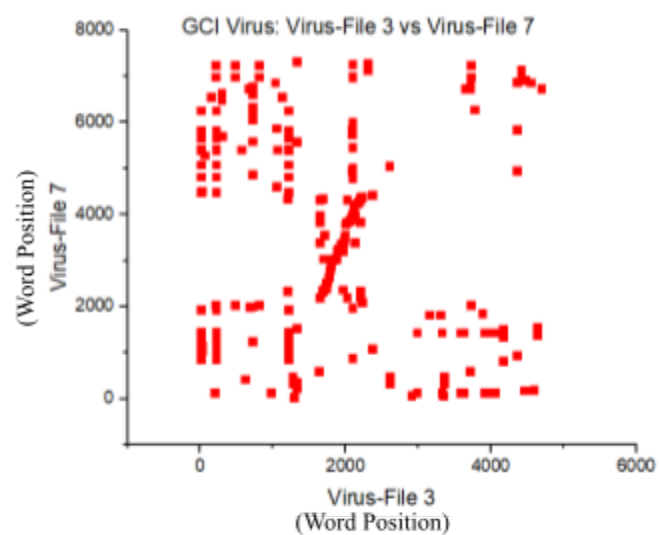
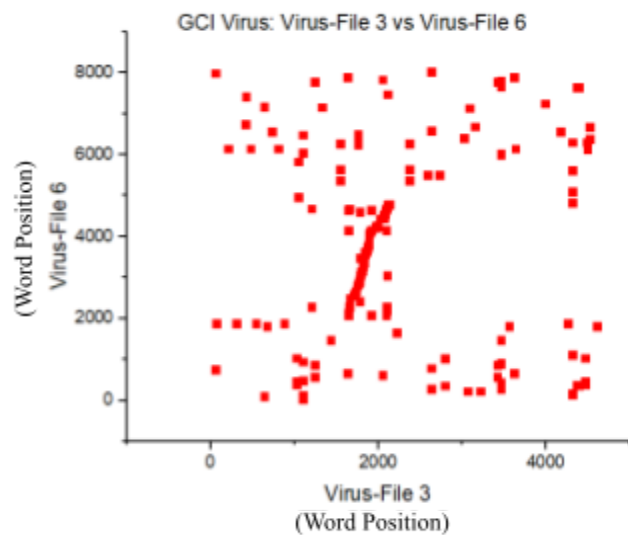
10.3. Graphs

10.3.1. Scatter Plots

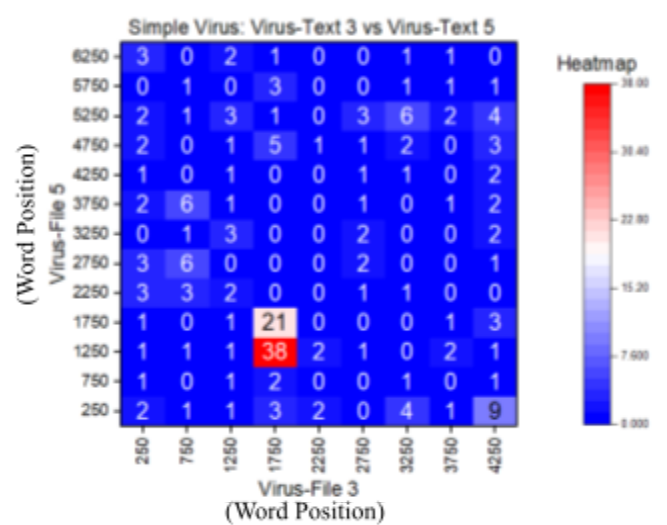
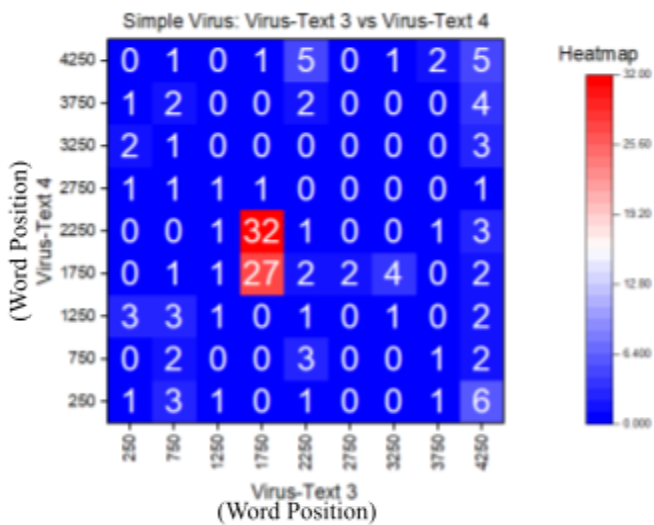
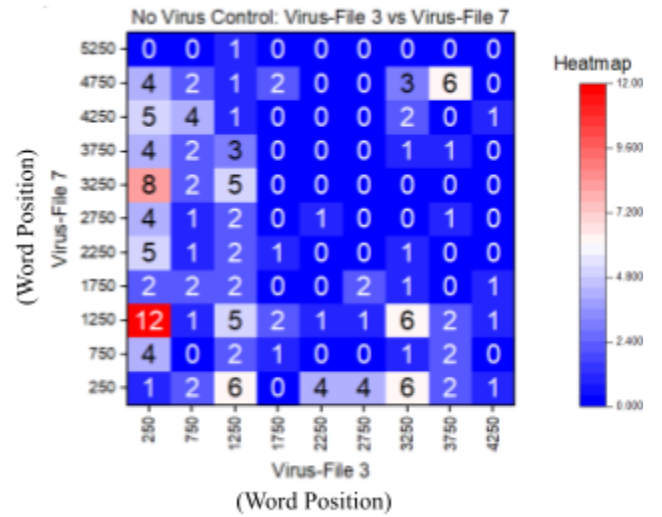
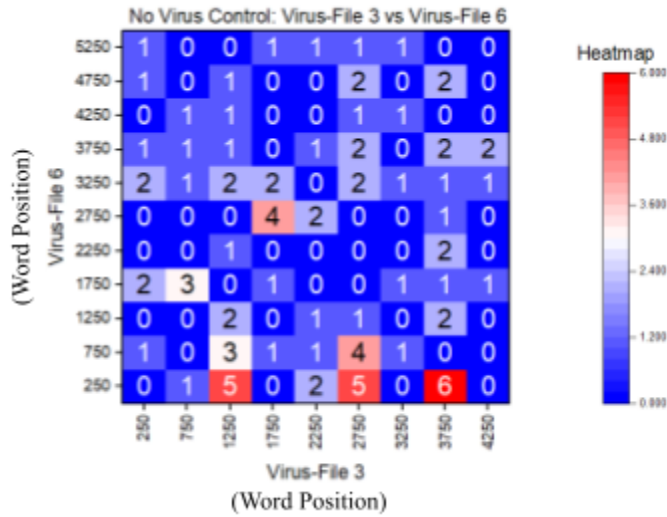
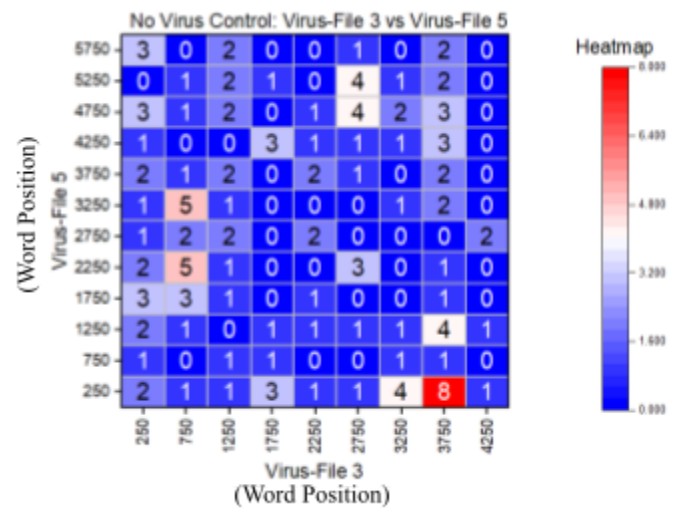
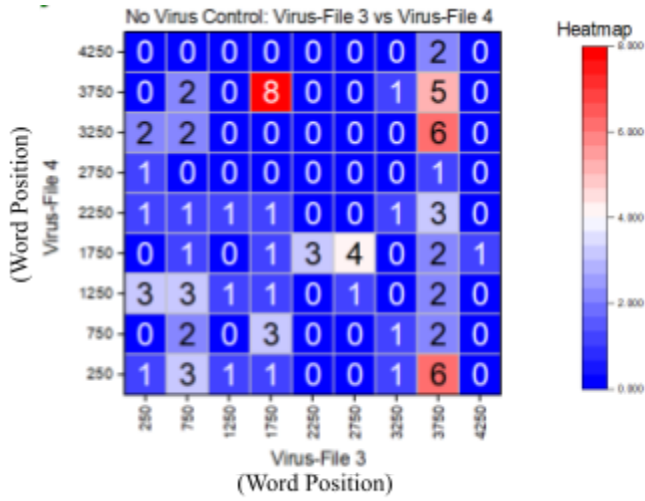


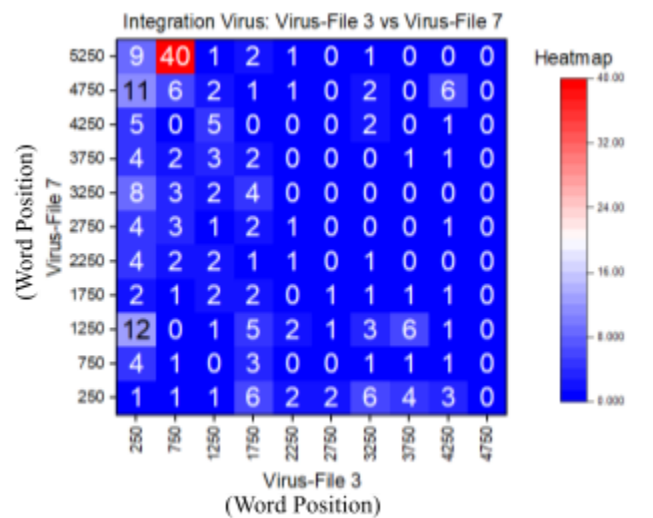
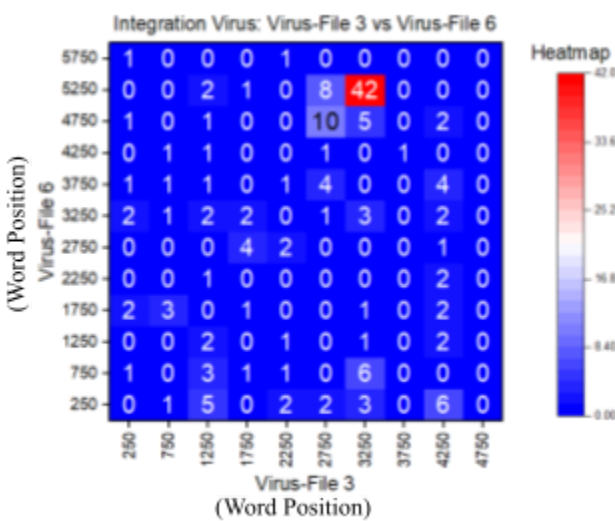
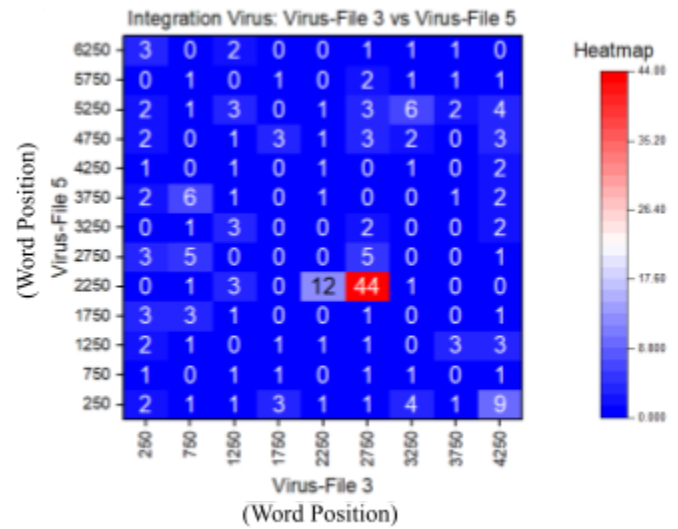
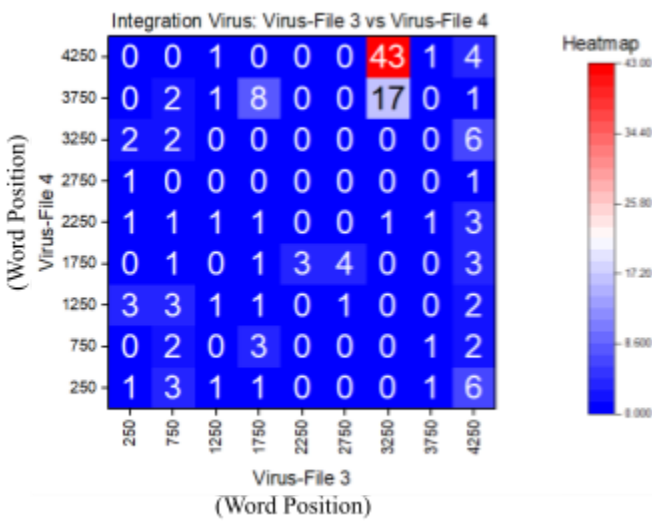
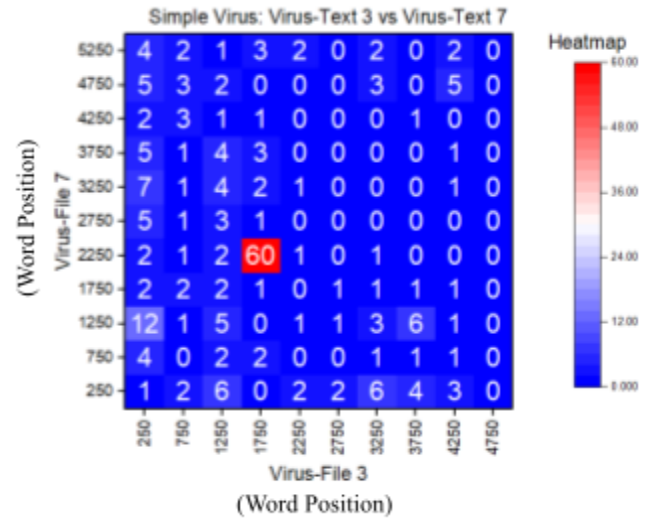
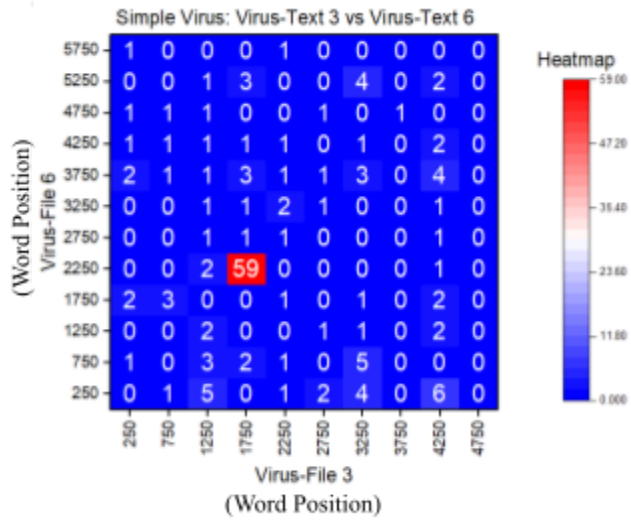


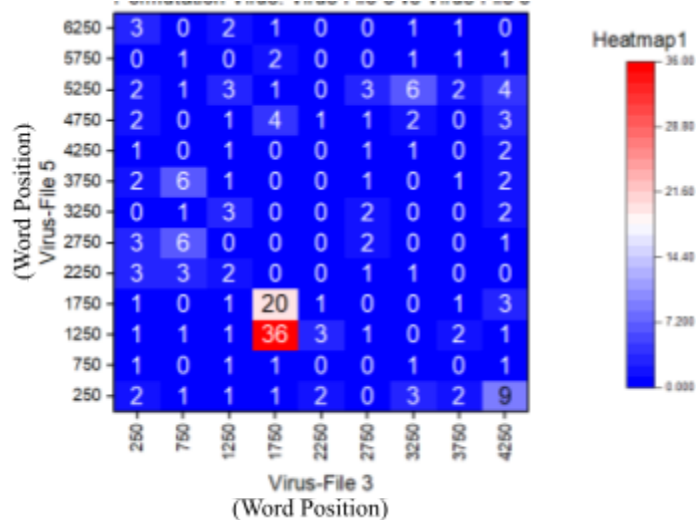
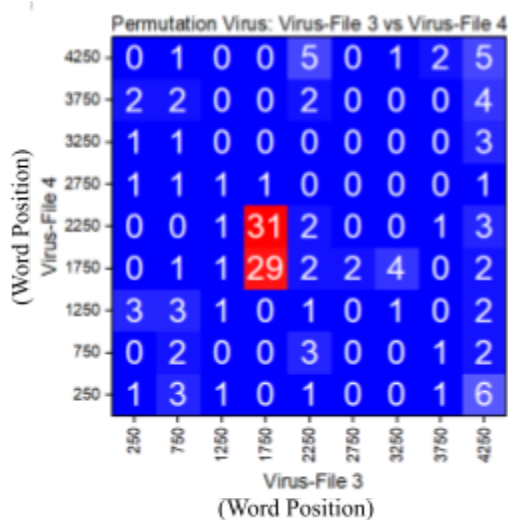
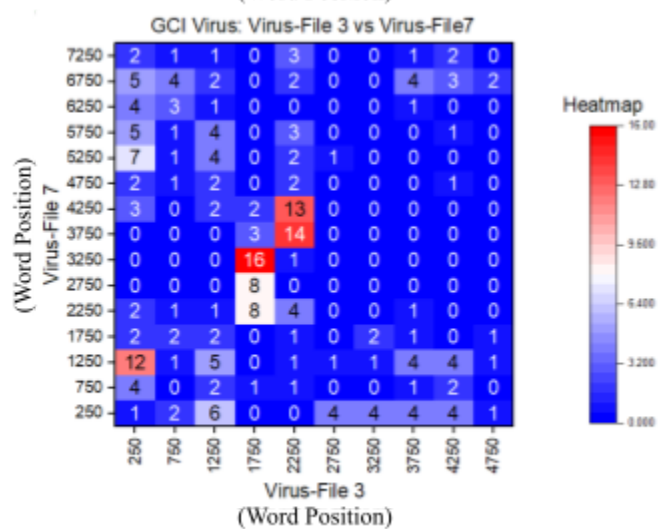
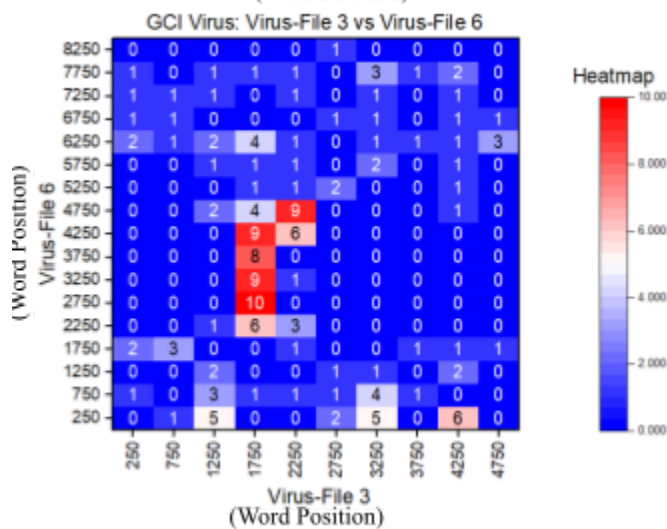
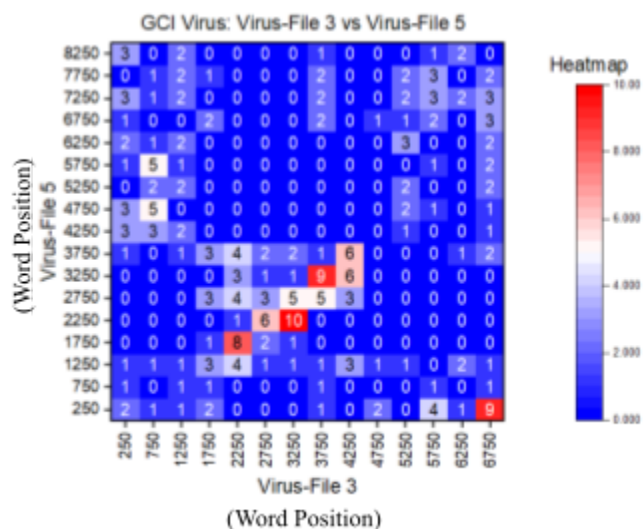
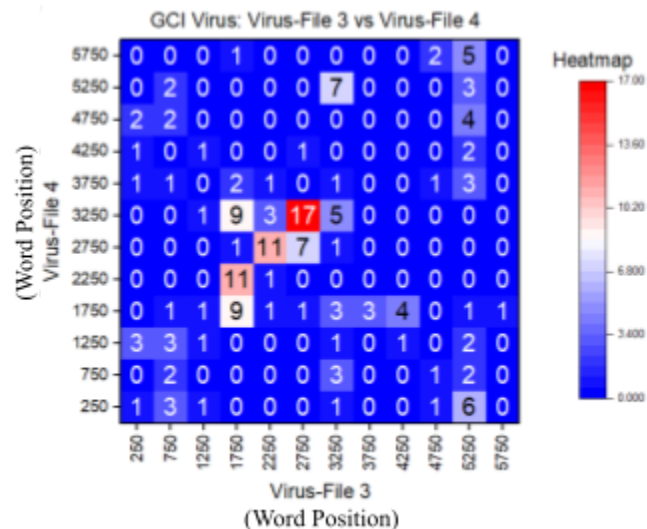


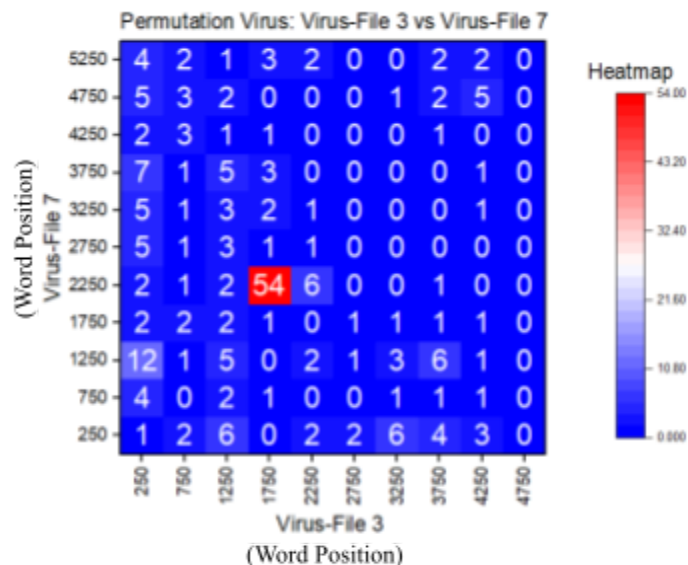
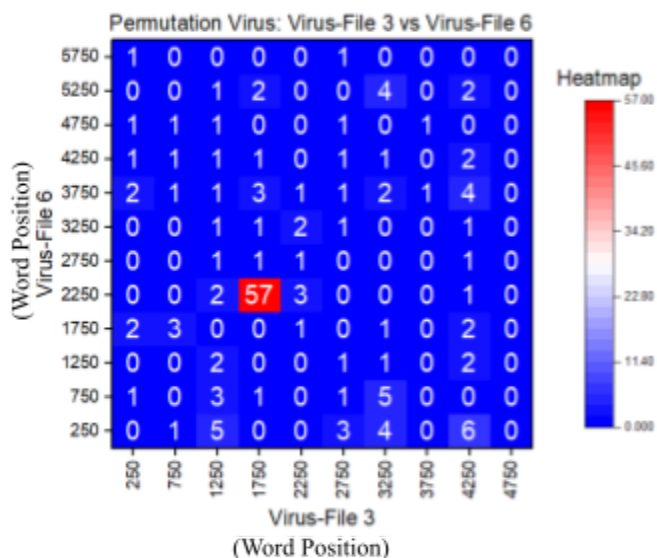


10.3.2. Heat Map









10.3.3. Enlarged Heat Map

