**Chapter 1**

**Introduction**

**Background of the Study**

In today’s generation, computer technology contribution had been known and experienced worldwide. Since the computer technology’s invention, the world has changed drastically.

Most people have been dependent on computer, it continually changed the way they live, and it rolled out to be very important in making their life more progressive through applying computer on how they work and how they deal with their everyday lives. Computer technology also contributed in improving the area of our education, industry, travel and health. Thus, as the rapid growth of population, computers also made its way for the highly-technological gadgets as the latest trend.

Computers caused a big impact in the society but due to the incredible growth of computers, risks and problems may arise. Technologies will always have a negative effect depends on how the technology is being used and who using it.

Invasion of privacy especially in written documents is one of the serious problems. As for students; individual confidential data, personal data, have the risks of invasion but due to the advancement of technology and through the knowledge of IT people problem like invasion of privacy may lessen up.

Encryption is the conversion of plaintext into ciphered text that is unreadable by unauthorized person unless they have the tool to decipher the text. Many kinds of encryption exist today for it is effective in hiding secrets most especially of a written text. With regards to this, recognizing of encrypted text has become rampant.

The researcher aimed to create a new and safe way of encrypting text. Encryption alone is not enough, seeing an encrypted text may confuse someone with its meaningless look and trigger someone’s curiosity about what it is and why does it looks like that. So in addition, the study will focus on encrypting the text not into its meaningless look but into a new word, phrase or sentence that is in a correct grammar that can’t make someone doubt about the text. Monte Carlo and NLP Algorithms are will be used in the study.

Thus, the study Grammar Encryption and Retrieval using Monte Carlo and NLP

Algorithm was conducted to develop a new way of encrypting and retrieving software application and bring encryption into its new level.

**Statement of Objectives**

The researcher aimed to develop a Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithms.

Specifically, this study is conducted with the following objectives:

1. accept a sentence, a phrase or a paragraph and accept a password;

2. encrypt text using Monte Carlo and NLP Algorithms;

3. retrieve text; and

4. evaluate system using ISO 25010.

**Conceptual Framework**

PROCESS

* Accept text(s)
* Accept password
* Tokenize text(s) into words
* Count number of words
* Encrypt input text by applying Monte Carlo Algorithm to randomly choose a substitute word
* Parse words
* Apply POS tagging to check if the output is syntactically correct.
* Retrieve the encrypted text using password

- Grammar check

-Decrypt ciphered text

INPUT

* Text in sentence, phrase or paragraph format
* Password

OUTPUT

Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm

SOFTWARE EVALUATION

ISO 25010

Figure 1 *Conceptual Framework of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm*

Figure 1 shows the conceptual framework of the Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm.

The inputs to the software application are text in a sentence, a phrase or a paragraph format; as well as password which will be used to retrieve the original text.

The processes of the system include accepting the text input, accepting the password, tokenizing the input to count the number of word(s), encrypting input text by applying Monte Carlo Algorithm to randomly choose a substitute word, parse words and apply POS tagging to check if the output is syntactically correct and retrieve the encrypted text using password.

The output system is the Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm is a grammatically correct encrypted text that can be evaluated using the ISO 25010 quality standards.

**Definition of Terms**

To easily understand the terms used in this study, the following are defined:

**Encrypt**. It is the process of converting a plaintext into an encrypted text. (OxfordDictionaries, n.d.).

In this study, encrypt refers to converting the original text into its new state by replacing it with new words that are syntactically correct.

**Grammar** . The structure of language in general, usually taken as consisting of

syntax, semantics, and morphology. It is the set of rules that explain how words are being used in a language. (Merriam-Webster, n.d.)

In this study, grammar will be use as a basis in checking the syntactic structure of the output of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm.

**Input**. The data gathered or accepted by computer. (ComputerHope, n.d.).

In this study, input refers to the text either word, sentence, phrase or a paragraph inputted by the user to be encrypted to its new state.

**Monte Carlo Algorithm.** Is a [randomized algorithm](https://en.wikipedia.org/wiki/Randomized_algorithm) whose output may being

correct with a certain [probability](https://en.wikipedia.org/wiki/Probability). It uses randomness and statistics as a result. (Wikipedia, n.d.).

In this study, Monte Carlo Algorithm will be use in randomly choosing a word to replace the original input.

**Natural Language Processing****Algorithm (NLP)**. Is a field of [computer science](https://en.wikipedia.org/wiki/Computer_science), [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) and [computational linguistics](https://en.wikipedia.org/wiki/Computational_linguistics) concerned with the interactions between computers and [human (natural) languages](https://en.wikipedia.org/wiki/Natural_language). (Algorithmia, n.d.)

In this study, NLP Algorithm will be use to check the output of the Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm syntactically.

**Parsing**. It is the process of analyzing a sentence in terms of grammatical constituents, identifying the part of speech, inflectional form and syntactic function. (Dictionary.com,n.d.)

In this study, parsing will be use to analyze if the output is grammatically correct

based on grammar rules.

**Part Of Speech (POS) Tagging.** It isthe task of determining what Part of Speech the word is. It defines what POS tags exist in the given like the most common; noun, verb, determiner, adjective and adverb. (Naber, 2003)

In this study POS tagging is used to define what part of speech is the random output words of the Monte Carlo Algorithm and if it match with the given format.

**Password**. It is a secret word or phrase that is used to gain admission to something. It is a string of characters that allows access to a computer, interface, or system. (Dictionary.com,n.d.)

In this study, the password refers to the input provided by the user that will be used for retrieving the encrypted text.

**Retrieve**. It is the process of getting or bringing back something to its former state. (Dictionary.com, n.d.).

In this study, retrieve refers to converting the encrypted text back to its original

state.

**Tokenization.** Given a character sequence and a defined document unit, tokenization is the task of chopping it up into pieces, called *tokens* , perhaps at the same time throwing away certain characters, such as punctuation. These tokens are often loosely referred to as terms or words, but it is sometimes important to make a type/token distinction. A token is an instance of a sequence of characters in some particular document that are grouped together as a useful semantic unit for processing.

(https://nlp.stanford.edu/IR-book/html/htmledition/tokenization-1.html)

**Grammar Encryption and Retrieval Using Monte Carlo and NLP Algorithms.** The system will encrypt the input text of the user and will output a grammatically correct encrypted text using Monte Carlo and NLP algorithm.

**Significance of the Study**

The result of the study will be beneficial to:

**Students**. The results of this study will be used to secure private written text of the students. It will help student to know how Monte Carlo Algorithm and NLP Algorithm are being applied in this study.

**Programmer**. The Monte Carlo Algorithm and NLP Algorithm will help the programmer to mold up their ideas to develop encryption/decryption in another way to achieve security of files.

**Private Individuals**. The results of this study will be used by somebody who wants to keep their information confidential.

**Future Researchers**. The study result will serve as a reference for the future researcher for their own study related to cryptography.

**Scope and Limitation of the Study**

This study entitled “Grammar Encryption and Retrieval Using Monte Carlo and NLP Algorithm” focuses on developing a new way of encrypting and retrieving text(s) using Monte Carlo and NLP Algorithm, and output a grammatically correct encrypted text. The system can encrypt text like word, phrase or even sentences and substitute it with a grammatically correct words randomly choose from the Dictionary database. Word and group of words in any languages not more than to 50 are acceptable in the system. This system will be design for standalone computers, and it will install in an offline environment. Text to be encrypted can be in any language, the substitute text would be in English language. Files in any format are not accepted in the system.

**Chapter 2**

**Review of Related Literature**

This chapter includes the related literature and studies that give ideas to the researchers in the formulation of the study.

**Encryption**

 It is the process of encoding a text in a way that only authorized parties can access it. It is the science of providing security for information. In an encryption scheme, the original text called [plaintext](https://en.wikipedia.org/wiki/Plaintext) is encrypted using an encryption algorithm, generating [an](https://en.wikipedia.org/wiki/Ciphertext) encrypted that can only be read if decrypted. Encryption generally divided into two categories namely Symmetric-key encryption and Public-key encryption. In symmetric-key the encryption and decryption keys are the same, the both parties have the same key for successful secure communication while in public-key encryption, the encryption key is available for anyone to use and encrypt messages. The receiver party only has the decryption key to turn the messages into readable (author, year).

In this study, the concept of symmetric-key encryption will be used in hiding the real content of the input text wherein the user will only have one key to start encrypting and in order to retrieve the encrypted text. Encryption will be done by replacing the input text of the user with new sets of words.

**Information Retrieval**

Information retrieval is an activity of getting [information](https://en.wikipedia.org/wiki/Information) resources relevant to an

information need from a collection of information resources. Searches can be based on [full-text](https://en.wikipedia.org/wiki/Full_text_search) or other content-based indexing. These automated information retrieval systems are used to reduce what has been called "[information overload](https://en.wikipedia.org/wiki/Information_overload)". Many [universities](https://en.wikipedia.org/wiki/University) and [public libraries](https://en.wikipedia.org/wiki/Public_library) use information retrieval systems to provide access to books, journals and other documents.  One of the most visible  [information retrieval applications](https://en.wikipedia.org/wiki/Information_retrieval_applications) are the [web search engines](https://en.wikipedia.org/wiki/Web_search_engine).

An information retrieval process begins when a user enters a query into the system. Queries are formal statements of information needs, for example search strings in web search engines. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of [relevancy](https://en.wikipedia.org/wiki/Relevance_(information_retrieval)). An object is an entity that is represented by information in a content collection or [database](https://en.wikipedia.org/wiki/Database). User queries are matched against the database information. However, as opposed to classical SQL queries of a database, in information retrieval the results returned may or may not match the query, so results are typically ranked. This ranking of results is a key difference of information retrieval searching compared to database searching. Depending on the [application](https://en.wikipedia.org/wiki/Information_retrieval_applications) the data objects may be, for example, text documents, images, audio, [mind maps](https://en.wikipedia.org/wiki/Mind_maps) or videos. Often the documents themselves are not kept or stored directly in the IR system, but are instead represented in the system by document surrogates or [metadata](https://en.wikipedia.org/wiki/Metadata) .Most IR systems compute a numeric score on how well each object in the database matches the query, and rank the objects according to this value. The top ranking objects are then shown to the user. The process may then be iterated if the user wishes to refine the query. (wikipedia.org, n.d.)

In this study, retrieval will be use as a counterpart of encryption wherein once the

User input the encrypted text and enters the password he/she provided during the encryption process, the original input that was saved in the database will be retrieve and be send back to the user.

**NLP Algorithm**

Natural Language Processing or known as NLP algorithm is the field of

understanding computer and manipulating the human language. Through NLP, computers can analyze, understand and derive the meaning of a human language in a useful way, it enables the computer to perform task such as text summarization, translation, named entity recognition, converting text to speech, relationship extraction, correcting grammar, sentiment analysis, speech recognition, and topic segmentation.

NLP algorithms are typically based on machine learning algorithm.

* Summarizer- use to extracts the most important blocks of text while neglecting the irrelevant text.
* Parsey MacParseface-it is a language parsing deep learning model created by Google that uses Point-of-Speech tagging.
* Named Entity Recognition-identify the type of entity that was extracted if it is a person, place or an organization.
* Sentiment Analysis- used to define the sentiment of a string of a text whether it is in negative state to neutral or to very positive.
* PorterStemmer- it reduces the words into its roots.
* Tokenizer- breaking up text into tokens.

Grammars and Parsing in NLP

Common denotations of Syntactic categories in NLP

* np - noun phrase
* vp - verb phrase
* s - sentence
* det - determiner (article)
* n - noun
* tv - transitive verb (takes an object)
* iv - intransitive verb
* prep - preposition
* pp - prepositional phrase
* adj – adjective

Context-Free Grammar is a list or rules that identify the set of all formed sentences in a language. The rules are composed of a left-hand side and a right-hand side. The left-hand side identifies the syntactic category while the right-hand side defines the alternative component parts reading from left to right.

Figure 2.0 below shows a simple CFG that describes the sentences from a small subset of English. CFG defined sentence in the language as a series of words derived by systematically applying the rules that begin on its left-hand side. Here is the example of the CFG:

A parse of a sentence is “the giraffe dreams” is:

s => np vp => det n vp => the n vp => the giraffe vp => the giraffe iv => the giraffe dreams

One of the best ways to describe a parse is through showing its parse tree. Parse of the sentence is a series of rule application in which a syntactic category is replaced by the right-hand side of a rule that has a category on its left-hand side, and the final rule application yields the sentence itself.

s np vp

np det n s

vp tv np np vp

iv det n iv

det the the giraffe dreams

a

an

n giraffe

apple

iv dreams

tv eats

dreams

Figure 2*.*1 *A Grammar and Parse Tree for “The Giraffe Dreams”.*

As shown in Figure 2.1 below is an example of an ambiguous grammar and a partial parse tree. The permit ambiguity is responsible of ensuring that a sentence has all its parses. (Algorithmia, 2016).

s np vp

np det n

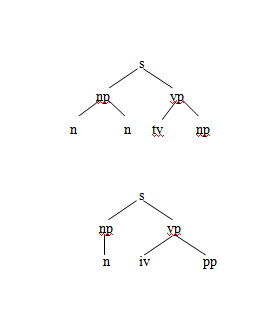
n

n n

vp tv np

iv pp

pp prep np

 det a

an

n fruit

apple

flies

iv flies

tv like

prep like

Figure 2.1 *An Ambiguous Grammar and Partial Parse Trees for “Fruit Flies Like an*

*Apple.”*

The researchers have studied the Natural Language Processing to apply the gained

ideas and methods in developing the Grammar Encryption and Retrieval using Monte

Carlo and NLP algorithm. In this study, NLP algorithm will be used for creating the grammar checker of the system to output a grammatically correct encrypted text.

**Monte Carlo Algorithm**

A Monte Carlo is a [randomized algorithm](https://en.wikipedia.org/wiki/Randomized_algorithm) whose output may be incorrect with a certain [probability](https://en.wikipedia.org/wiki/Probability). It lets you see all the possible outcomes of your decisions and assess the impact of risk, allowing for better decision making under uncertainty. Monte Carlo simulation is a computerized mathematical technique that allows people to account for risk in quantitative analysis and decision making. Monte Carlo simulation risk analysis by building models possible results by substituting a range of values for any factor that has inherent uncertainty. It then calculates results over and over, each time using a different set of random values from the probability functions. Monte Carlo produces distributions of possible outcome value.

By using probability distributions, variables can have different probabilities of different outcomes occurring. Probability distributions are a much more realistic way of describing uncertainty in variables of a risk analysis. Common probability distributions include:

* Normal – Or “bell curve.” The user simply defines the mean or expected value and a standard deviation to describe the variation about the mean. Values in the middle near the mean are most likely to occur. It is symmetric and describes many natural phenomena such as people’s heights. Examples of variables described by normal distributions include inflation rates and energy prices.
* Lognormal – Values are positively skewed, not symmetric like a normal distribution. It is used to represent values that don’t go below zero but have unlimited positive potential. Examples of variables described by lognormal distributions include real estate property values, stock prices, and oil reserves.
* Uniform – All values have an equal chance of occurring, and the user simply defines the minimum and maximum. Examples of variables that could be uniformly distributed include manufacturing costs or future sales revenues for a new product.
* Triangular – The user defines the minimum, most likely, and maximum values. Values around the most likely are more likely to occur. Variables that could be described by a triangular distribution include past sales history per unit of time and inventory levels.
* PERT- The user defines the minimum, most likely, and maximum values, just like the triangular distribution. Values around the most likely are more likely to occur. However values between the most likely and extremes are more likely to occur than the triangular; that is, the extremes are not as emphasized. An example of the use of a PERT distribution is to describe the duration of a task in a project management model.
* Discrete – The user defines specific values that may occur and the likelihood of each. An example might be the results of a lawsuit: 20% chance of positive verdict, 30% change of negative verdict, 40% chance of settlement, and 10% chance of mistrial.

During a Monte Carlo simulation, values are sampled at random from the input probability distributions. Each set of samples is called iteration, and the resulting outcome from that sample is recorded. Monte Carlo simulation does this hundreds or thousands of times, and the result is a probability distribution of possible outcomes. In this way, Monte Carlo simulation provides a much more comprehensive view of what may happen. It tells you not only what could happen, but how likely it is to happen. (Palisade, 2017).

In this study, Monte Carlo Algorithm will be used to randomly pick up new words to replace the input.

**ISO 25010 Standard**

The ISO 25010 standard is the basis of the product quality evaluation system and determines which quality of characteristics will be taken when evaluating the system. The quality of the system is the degree to which the system satisfies and implies the needs of the stakeholders. The ISO/IEC 25010 consist of eight quality characteristics the functional suitability, performance efficiency, portability, compatibility, maintainability, usability, reliability, and security. In terms of functionality, the system will be tested through its functional completeness, functional correctness, and functional appropriateness. The functions are those that satisfy or implied the needs of the system. The performance efficiency, this attribute represents on the relationship between the level of performance of the system and amount of resources used under stated conditions. Portability this refers to the capability of the system to be transferred from one usage environment to another. The system will be evaluated through its installability, adaptability and replaceability. Maintainability is an attribute that refers to the capability to modified system to improve it, correct it or adapt its changes. This will be tested through its modularity, reusability, analyzability, modifiability, and testability. Usability this refers to the capability of the system that can provide less effort for the users to use, whether the system achieved the goals of effectiveness, efficiency, and satisfaction for the user. Usability was determined through its operability, appropriateness recognizability, learnability, accessibility, user error protection, and user interface aesthetics. Compatibility bear on the capability of the system that can perform exchange information with other system. Reliability bear on the capability of the system to maintain its level of performance under specified conditions for a specified period of time and effort must needed. Reliability will be determined through its maturity, availability, fault tolerance, and recoverability. Security refers to the ability of the system to protect the information or the data of the users and to ensure that the data can accessible only for the authorized person to have access. Security will be determined through its confidentiality, integrity, non-repudiation, accountability, authenticity.

(iso25000.com,2015).

The ISO 25010 standard will be the basis in evaluating the performance of the system.

**Crypten : Encryption**

Crypten : Encryptionis an application created by Dynamitechetan wherein user can encrypt anything in their mobile.

The application is a secure way to keep text or files with confidential data encrypted using most secure encrypting and hashing algorithms. (Dynamitechetan, n.d.).

Using the latest material design principles, Crypten: Encryption provides an easy way to

keep files and text messages secret and secure.

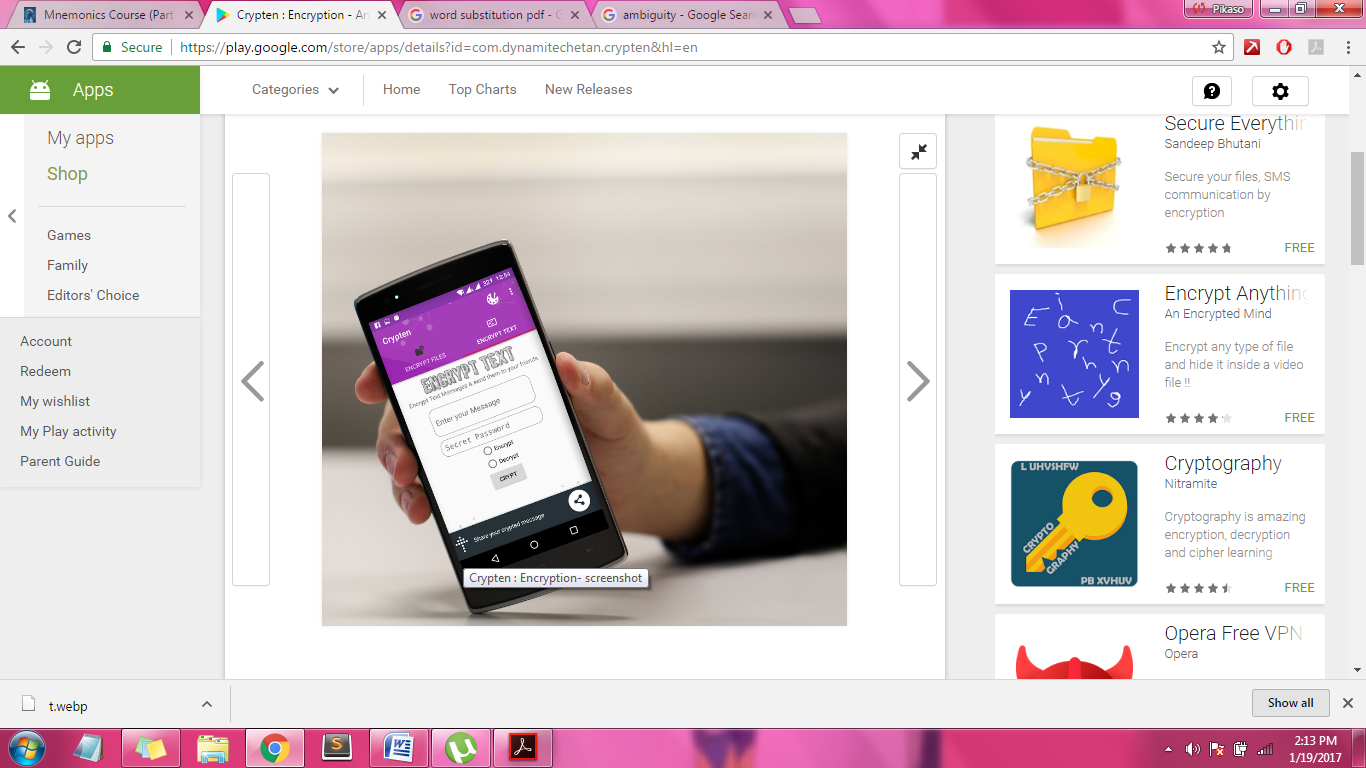


Figure 3.*0 Screenshot of Crypten:Encryption Application*

Figure 3 shows the screenshot of the Crypten:Encryption Application. The file can be encrypted with a text-based password or select a file as password. Text messages are encrypted using AES algorithm which is protected by a password. These encrypted messages can be sent to messaging apps and social networks. The application also enables the user to send secret information like bank account numbers and license information to any person ( author, year).

In this study, the researcher relate to Crypten:Encryption in which in the application the user need to provide a password to perform the encrypting and decrypting of the input text but unlike the Crypten: Encryption, the researchers study entitled Grammar Encryption and Retrieval focuses only on encrypting text and files are not included. It aims to bring encryption into its new level.

**Crypditor**

Figure 4.0 below shows the screenshot of the **Crypditor**. It’s a software that enables a free text editor to encrypt a text with AES algorithm. The **Crypditor** works like a text editor wherein the user can enter a text in the provided text box and save it by providing a password. Crypditor can import any text file to encrypt. The user can open multiple files in its tabbed windows. It creates executable of the encrypted text and require no other software to decrypt it.

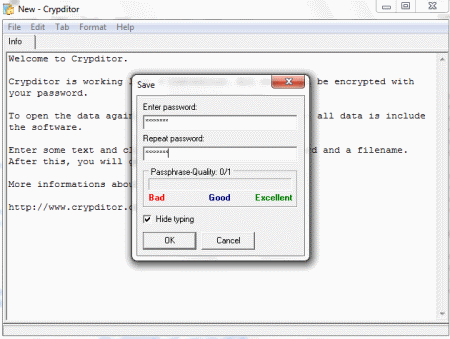


Figure 4.0 *Screenshot of Crypditor*

The data file are can be copied on a flash drive. The user doesn’t need to install any software: the encrypted data is hidden within the software and if it is lost, no one can open it text file because it is protected by the password. **Crypditor** does not require any installation.

In this study, the researchers relate the Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm wherein it will also accept a password from the user.

**SecretPad**

SecretPad is a simple plain text encryption software which encrypts the text using your password. It has simple interface similar to Windows Notepad. It can be used for storing personal records. You can set the password by pressing Ctrl + W or choosing Password from file menu. It saves the file in CDT format. To decrypt the file you must provide the right password to SecretPad. It requires no installation.

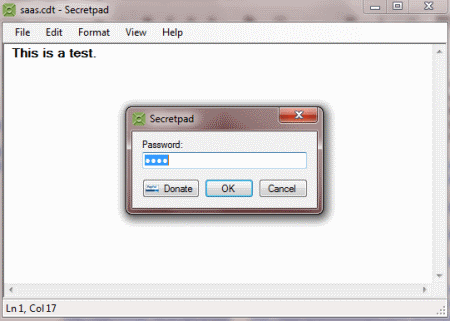


Figure 5*.0 Graphical User Interface of SecretPad*

Figure 5.0 shows the GUI of SecretPad. From this software, the researchers got the idea of creating the GUI of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm user friendly and easy to understand.

**Chapter 3**

**STATEMENT OF PROJECT ACTIVITIES**

**Project Description**

Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm is a system that brought encryption into its new level. Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm is created especially to the user who wants to protect the confidentiality of their written data.

The process of encryption is known to encrypt text and turn it into its unreadable state to hide the real message or content of the text. In today’s generation, the same idea of the process of cryptography exists. The researchers study focuses on encrypting the text but not just into its unreadable state but to its new look and group of words it composed based on the English Dictionary.

The user will input a plaintext in the text box provided in the system. Once the user click the encrypt button, the user will then be ask to enter a password that will be use in decrypting back the encrypted text. After the user provided a password the cipher text will then be display. In the retrieval process, the user will enter the cipher text and the password he/she provided and the original plaintext will be display. The output of the system are can be copy by the user by clicking the copy button and can paste it in any document like Microsoft word or Notepad where the user want to save a copy of the encrypted text. The researchers conducted an intensive research study about the system and about the Monte Carlo and NLP Algorithm that will be used in creating this grammar encryption and retrieval system.

**Project Activities**

The methodology of this system development is to have a research regarding on cryptography of texts such as word, phrase and sentence as a result of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm.

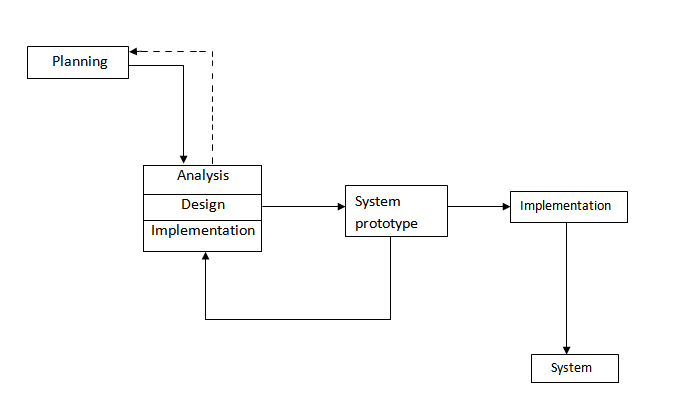


Figure 6.0 *Prototyping-base Methodology*

Figure 6.0 shows the Prototyping-base Methodology that the researchers will use to develop the Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm. This type of SDLC model is appropriate in developing the system because it uses system prototype to refine such key issues before the system will be deploy. In this methodology, the fundamentals of analysis and design were executed on a system prototype.

The Prototype model begins with planning. In this stage the researchers collected and study systems to gain idea, the researchers relate on what kind of system they need to develop. In this stage, the fundamental process of understanding what kind of system are built; the researchers identify the problem why they need to develop this system, the researchers research about related studies that also performs encryption and retrieval process. The researchers discovered that the encrypted texts are easy to recognize by someone because of its unusual look. From this problem, the researchers came up with the idea of creating the Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithms. This system performs encryption in readable state with correct grammar.

Analysis was the second stage. The researchers planned a solution on how to develop the system. The researchers analyzed and determined the algorithm that will be used to develop the system. In this stage, the researchers created a documentation to make the system clearer during analysis. The diagrams used in developing system and pseudo code were taken into consideration.

On the design stage, the researchers determined the platform of the system. The researchers designed the user interfaces, and the design of the desired output. The User Interface was designed to be user friendly that easy for the users to understand. The System Prototype was developed during the implementation stage. The users provided comments which were used to reanalyze, redesign and re-implement of the system. After the system prototype approved, the researchers make the actual design of the system. The process continued until the researchers will reach the functionality of the system whether the system operated properly. Continuous refinements and observation by the developers before the system was deployed.

**Requirement Analysis**

To fully understand the flow of the system, the Context Diagram and Diagram 0 are presented in this chapter.

Text

USER

Grammatically correct Encrypted Text

Password

0

Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm

Grammatically correct Encrypted Text

Text

*Figure 7.0 Context Diagram of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm*

Figure 7.0 shows the Context Diagram of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm. The User entity will provide the input text and will click the Encrypt button to start the encryption of the text. In the encryption, the User will provide a Password that will be used in the retrieval process. The user will receive the encrypted text. In the retrieval process, the user will enter the encrypted text, click the Retrieve button and will enter the Password. The user will receive the original text. The User can click the Clear button to clear the display of the text box.

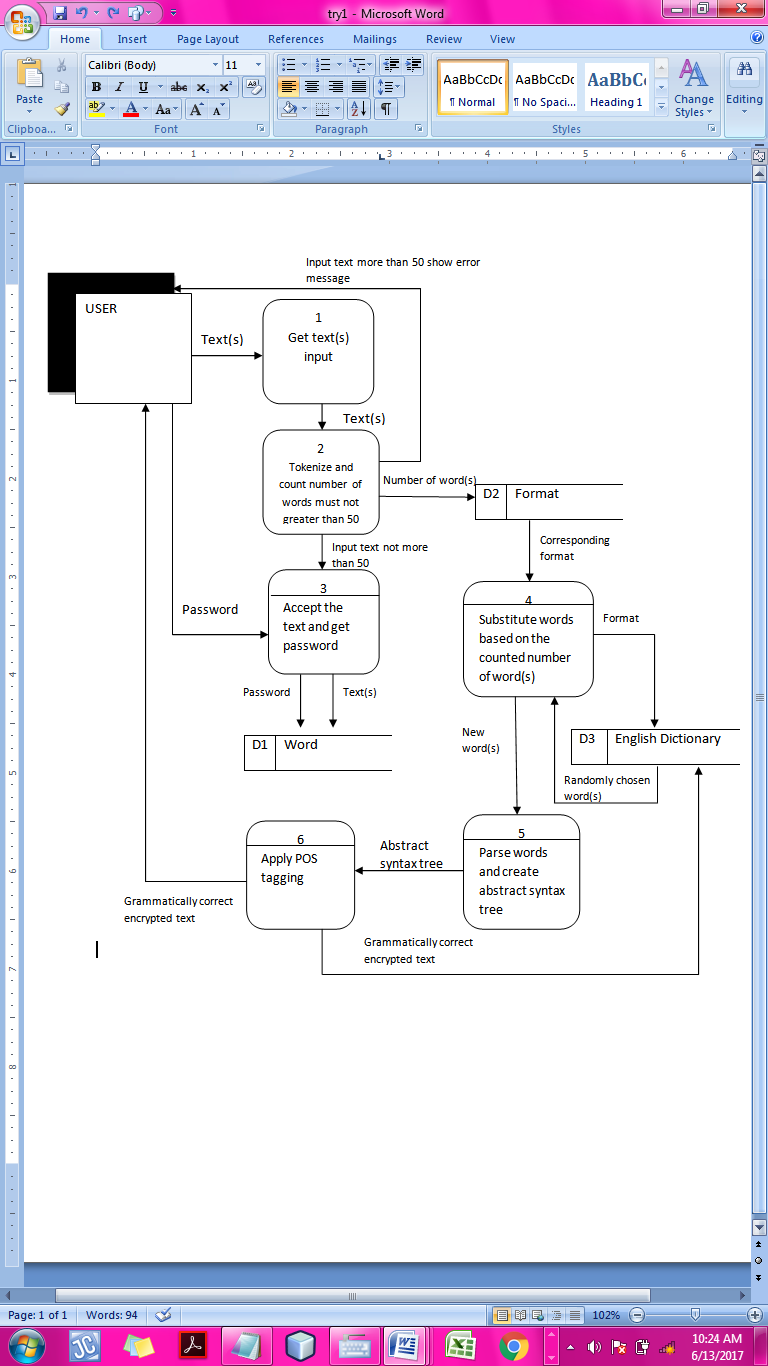
****

Figure 7.1 *Diagram 0 DFD of Grammar Encryption and Retrieval using Monte Carlo*

*and NLP Algorithm for Encryption Process*

Figure 7.1 shows the Diagram 0 DFD of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm for encryption process. In encrypting the text, the User will enter a text; the software application will tokenize the text into word(s) and count the number of word(s). If the number of word(s) is more than 50, it will send an error message to the user else the software application will accept the text and will prompt a password from the user and save it in Word Database

Password that will be use in decrypting the ciphered text. The User will input a text that will be saved in the Words Database. The system will count the numbers of words; based on the number of the words the system will determine the format that will get from the Format database. Monte Carlo will be applied in replacing the original words with new words from English Dictionary database. NLP Algorithm will be applying to check if the output is in correct grammar. Grammatically correct encrypted text will be sent to the user.

USER

1

Count number of words

Text

Text

Password

Password

|  |  |
| --- | --- |
| D1 | Words Database |

Number of words

Format ID

2

Determine the format

|  |  |
| --- | --- |
| D2 | Format |

Grammatically Correct Encrypted Text

Text format

3

Apply Monte Carlo

Substitute words

Format

New words

Wrong Grammar

|  |  |
| --- | --- |
| D3 | English Dictionary |

5

Apply NLP Algorithm

**w**

Figure 7.1 *Diagram 0 DFD of Grammar Encryption and Retrieval using Monte Carlo*

*and NLP Algorithm for Encryption Process*

Figure 7.1 shows the Diagram 0 DFD of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm for encryption process. In encrypting the text, the User will provide a Password that will be used in decrypting the cipher text. The User will input a text that will be saved in the Words Database. The system will count the numbers of words; based on the number of the words the system will determine the format that will get from the Format database. Monte Carlo will be applied in replacing the original words with new words from English Dictionary database. NLP Algorithm will be applying to check if the output is in correct grammar. Grammatically correct encrypted text will be sent to the user.

Success: password

User

1

Validate by matching input

password

Grammatically Correct Encrypted Text

|  |  |
| --- | --- |
| D1 | Words Database |

2

Retrieve Original text

Original Text

Text

Figure 7.2 *Diagram 0 DFD of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm for Retrieval Process*

Figure 7.2 shows the Diagram 0 DFD of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm for retrieval process. The User will input the grammatically correct encrypted text and the key code that he/she provided during the encryption process. The system will accept the encrypted text and the key code and will then retrieve the encrypted text. The original text output will be sent to user.

**Pseudo code of the Encryption process**

PROMPT sentence

GET sentence

Split sentence into words

GET tokenized words

COUNT words=words\_number

IF( words\_number!>=100)

{

RETURN error message

}

ELSE

{

ACCEPT sentence

PROMPT Password

GET Password

GET words\_number

COMPARE

GET text\_format

SELECT words randomly

READ Selected\_words

CHECK grammar\_of\_new\_words

IF(Selected\_words=text\_format)

{

DISPLAY the Grammatically Correct Encrypted Text

}

ELSE

{

PERFORM to Monte Carlo Process

}

}

END

DISPLAY Selected words

Figure 7.3 *Pseudo code of the Encryption process*

Figure 7.3. shows the pseudo code of the Encryption process. The user will input a text. If the text is equal or less than to 50, the system will accept the text. To start the encryption process the user need to provide a password. By applying Monte Carlo Algorithm the system will randomly replace the original input text with new text. The output will be check by NLP Algorithm by first parsing the words and by POS tagging it to check if it is syntactically correct. If the output is correct, the system will display the grammatically correct encrypted text if not the system will choose again another new set of words.

**Pseudo code of the Password**

PROMPT Password

GET Password

ACCEPT Password

END of input Password ()

Figure 7.4 *Input Password*

Figure 7.4 shows the input of password process. In order to encrypt a text the user must provide a password that will be save in the Words Database together with the input text of the user. This password will also be use to begin the retrieval process.

**Pseudo code of checking the grammar of the New Words**

GET new\_words

PARSE new\_words

CREATE abstract\_syntax\_tree

LABEL abstract\_syntax\_tree

GET labelled\_abstract\_syntax\_tree

GENERATE new\_words\_format

IF (new\_words\_format!= Text\_format )

{

PERFORM Monte Carlo Algorithm

}

ELSE {

RETURN sentence

}

DISPLAY grammar\_checked\_sentence

END of grammar\_check

Figure 7.5 *Pseudo code of Checking the Grammar of the New Words process*

Figure 7.5shows the pseudo code of checking the grammar of the new chosen words. Parsing will be applied to the new set of words to get its abstract syntax tree then the POS tagging will be apply to label the part of speech of the new set of words and if it matches with the saved text format in database the output will be displayed to user but if it doesn’t match, the system will choose another new set of words.

**Pseudo code of choosing new words**

CALL text\_format

for all part of speech (x)

LOOP at xi

GENERATE randInt

SELECT id from table

Order by RAND()

LIMIT(1)

IF randInt exist

Pull out from the table

ELSE

SELECT again

DISPLAY selected words

END

Figure 7.6 *Pseudo code of choosing new words*

Figure 7.6 shows the pseudo code of choosing new words. Based on the number of words that the user input, the system will call the text format. The text formats are composed of different part of speech that labeled as x. Each part of speech is labeled as xi. The system will loop to find the part of speech that consist the text format. If the particular part of speech exist, the system will randomly pull-out one id number that stands for the word until it satisfied the text format else will select again.

**Pseudo code of the Retrieval process**

PROMPT encrypted text

GET encrypted text

PROMPT Password

MATCH Password

CALL original text

DISPLAY original text

Figure 7.7 *Pseudo code of the Retrieval process*

Figure 7.7 shows the pseudo code of the Retrieval process. The user will enter

the password he/she provided during the encryption process. The input password will be

match in the Words Database and the original text will be sent to the user.

**Design Specification**

Graphical User Interface of Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm

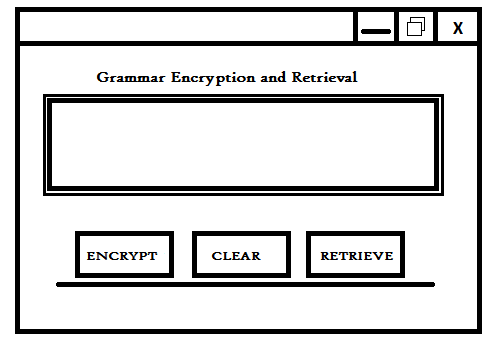


Figure 8 *Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithms Interface*

Figure 8 shows the main interface of the Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm. The user will input a text in the text box provided above the buttons and will click the encrypt button to start encrypting.

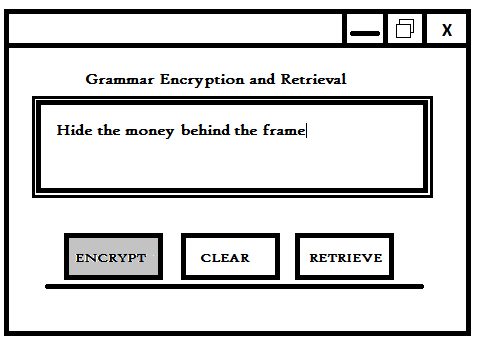


Figure 8.*1 Clicking Encrypt button*

Figure 8.1 shows the user inputted a text in the text box and click the encrypt button.

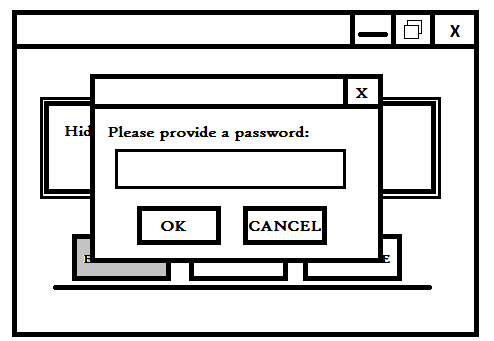


Figure 8.2 *Providing a Password*

Figure 8.2 shows the Password pop up box .Once the user click the encrypt button, the system will prompt a password from the user that will be use in decrypting back the cipher text.

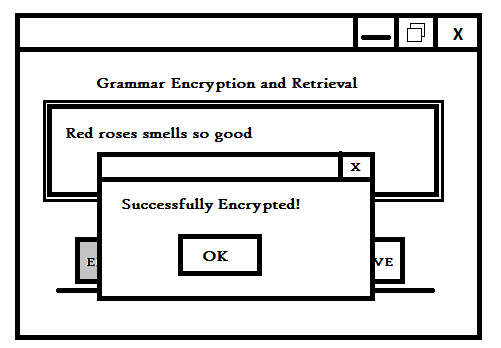


Figure 8.3 *Successfully Encrypted*

Figure 8.3. shows that the encryption process was successfully encrypted. Once the user click the ok button of the password box, the system will sent a message confriming that the encryption of the text was successful.

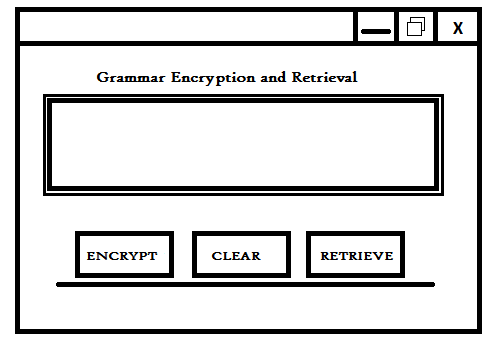


Figure 8.4  *Grammar Encryption and Retrieval using Monte Carlo and NLP*

*Algorithms Interface*

Figure 8.4 shows the shows the main interface of the Grammar Encryption and Retrieval using Monte Carlo and NLP Algorithm. The user will click the retrieve button to start the retrieval of the encrypted text.

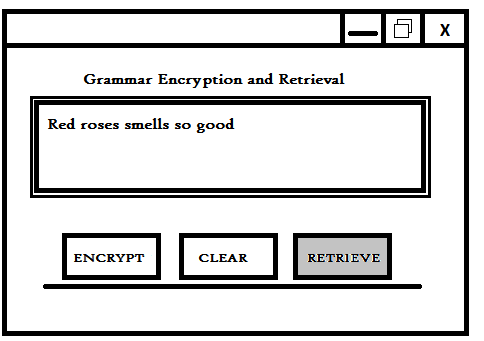


Figure 8.5 *Clicking Retrieve Button*

Figure 8.5 shows the retrieval process. The user will enter the encrypted text in the text box and click the retrieve button to retrieve the original input of the user.

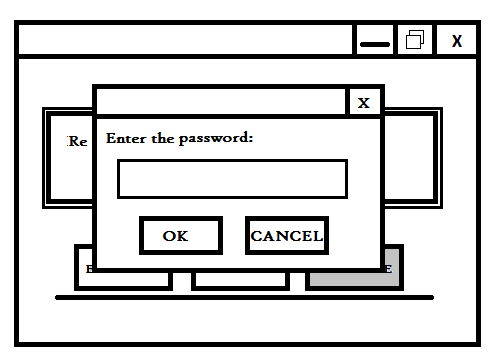


Figure 8.6 *Entering a Password*

Figure 8.6 shows the password pop up box asking for the user to enter the same password they enter in encrypting the text in order to decrypt the input of the user.

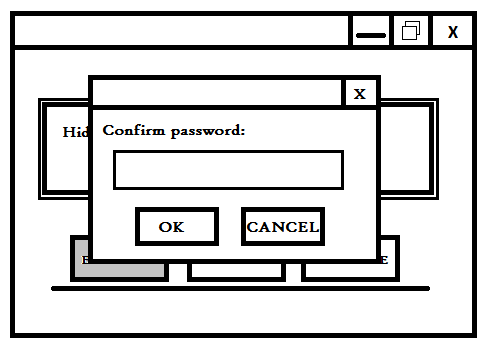


Figure 8.7 *Confirm a Password*

Figure 8.7 shows the password pop up box asking for the user to confirm the password he/she enter in order for the user to remember it .

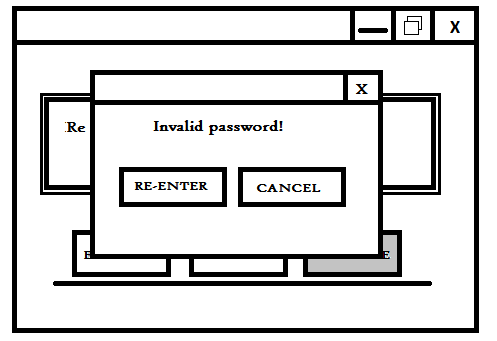


Figure 8.8 *Invalid Password*

Figure 8.8 shows the invalid password pop-up box sent by the system once the user enter a wrong password. The user can either cancel the decryption or re-enter the password again.

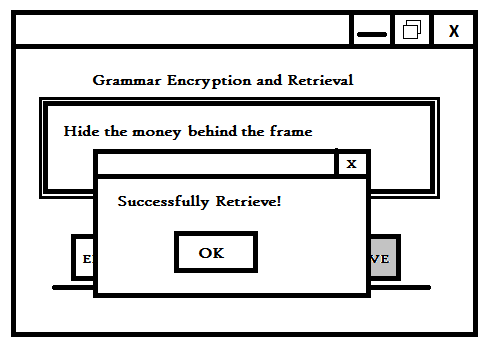


Figure 8.9 *Successfully Decrypted*

Figure 8.9 shows that the decryption process was successfull. Once the user enter the correct password, the cipher text will be converted back to its original state.

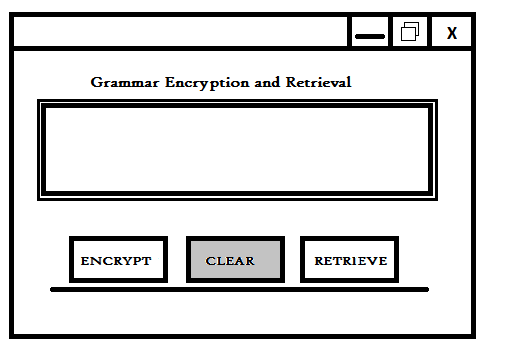


Figure 8.10 *Clear Process*

Figure 8.10 shows the clearing of the text shown in the text box. Once the user click the clear button text will be clear and the user can input a new text..

**Environment**

As the researchers conducted the study, the following were the different material used to obtain the study’s objectives.

The researchers used the Java Programming Language in implementing the Cryptography System with Grammar Checker using Monte Carlo and NLP Algorithm. Windows 7 Ultimate version 2009 Service Pack 1 Operating System installed in the System Unit. It has an Intel ® Celeron ® CPU 1005 M. Processor 1.90 GHz @ 1.90GHz speed with 2.00 GB (memory size) of RAM (Random Access Memory). The researchers tested, evaluated and implemented the Cryptography System with Grammar Checker using Monte Carlo and NLP Algorithm.

In addition, the researchers used Microsoft Word 2007 for creating, editing and documenting the study. Furthermore, the researchers will use 50k1.6.0-20 accompanied with Netbeans for the implementation or Cryptography System with Grammar Checker using Monte Carlo and NLP Algorithm.

**Testing and Operation**

White box and black box testing are essential in the development of the system. In this study the researchers system will be tested using the black box testing.

Black box testing is a technique that enabled the researcher to derive a set of input condition that will fully exercise all the functional requirements of the program. Black box testing is the process in which the tester is oblivious to the system architecture and does not have access to the source code. It means that a tester interact with the system user interface by providing inputs and examining outputs without knowing how and where the inputs are worked upon this technique. The possible user that will test the researchers system might include students of Iloilo Science and Technology.

The standard ISO 25010 will also be use as a basis of evaluating the performance of the Cryptography System with Grammar Checker using Monte Carlo and NLP Algorithm in terms of functional suitability, performance efficiency, portability, compatibility, maintainability, usability, reliability, and security

Both the Black box testing and the ISO 25010 standard will be the basis for the researchers in the development of Cryptography System with Grammar Checker using Monte Carlo and NLP Algorithm and preventing bugs and errors from occurring in the future.

**Project Evaluation**

The ISO 25010 standard is the basis of evaluating the performance of the system in terms of functional suitability, performance efficiency, portability, compatibility, maintainability, usability, reliability, and security.

The table below shows the level of satisfactory of the Cryptography System with Grammar Checker using Monte Carlo and NLP Algorithm will be evaluated by the respondent using the Likehert’z Rating Scale.

Table 1.

Likert’z Rating Scale for Evaluation

|  |  |
| --- | --- |
| Rate | Interpretation |
| 4.21-5.00 | Very Satisfactory |
| 3.41-4.20 | Satisfactory |
| 2.61-3.40 | Good |
| 1.81-2.60 | Fair |
| 1.00-1.80 | Poor |

Table 1 shows the Likert’z Rating Scale for Evaluation. The lowest rate is 1 which is interpreted as Poor and the highest rate is 5 that represent the Very Satisfactory. The average score will be taken by adding up the set of score and divide them by the number of score in the set. Based on evaluation, it will help the researchers to know if they accomplish their goals.