TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN CẤU TRÚC RỜI RẠC**

**MIDTERM ESSAY**

**Applied Probability and Statistics for IT**

*Người hướng dẫn*: **TRẦN LƯƠNG QUỐC ĐẠI**

*Người thực hiện*: **NGUYỄN THÀNH LỘC – 52000848**

Lớp **: 20050401**

Khoá  **: 24**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2021**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN CẤU TRÚC RỜI RẠC**

**MIDTERM ESSAY**

**Applied Probability and Statistics for IT**

*Người hướng dẫn*: **TRẦN LƯƠNG QUỐC ĐẠI**

*Người thực hiện*: **NGUYỄN THÀNH LỘC – 52000848**

Lớp **: 20050401**

Khoá  **: 24**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2021**LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành và sâu sắc nhất đến với các thầy cô của trường đại học Tôn Đức Thắng, đặc biệt là các thầy cô khoa Công nghệ thông tin môn Xác suất thống kê của trường đã tạo điều kiện cho em học tập ở trường cũng như giảng dạy online để có các kiến thức bổ ích giúp em làm các bài kiểm tra cũng như làm bài báo cáo một cách chỉnh chu nhất. Và em cũng xin chân thành cám ơn thầy Trần Lương Quốc Đại và cô Nguyễn Thị Huỳnh Trâm, thầy và cô đã nhiệt tình hướng dẫn hướng dẫn em hoàn thành tốt môn học.

Trong quá trình học tập, cũng như là trong quá trình làm bài báo cáo, có thể em còn nhiều sai sót, rất mong các thầy, cô bỏ qua. Đồng thời do trình độ lý luận cũng như kinh nghiệm thực tiễn còn hạn chế nên bài báo cáo không thể tránh khỏi những sơ xuất, em rất mong nhận được ý kiến đóng góp thầy, cô để em học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt hơn bài báo cáo sắp tới.

Em xin chân thành cảm ơn!

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm chúng tôi và được sự hướng dẫn của thầy Nguyễn Huỳnh Minh Duy;. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 21 tháng 04 năm 2022*

*Tác giả*

*A picture containing diagram

Description automatically generated*

*Nguyễn Thành Lộc*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

TÓM TẮT

Nhìn chung, đề bài báo cáo là dạng đề bài tiêu chuẩn dành cho mọi mọi sinh viên đang theo học môn Xác Suất Thống Kê. Đề bài có sự phân chia rõ rệt cho từng yêu cầu đề bài, đòi hỏi sinh viên phải thật sự nắm rõ lý thuyết và kiến thức kinh nghiệm trong quá trình thực hành để có thể áp dụng một cách chính xác nhất

Nội dung báo cáo gồm 4 chương:

- Chương 1: Khái niệm cơ bản về mã hoá, giải mã, mã hoá đối xứng và không đối xứng:

- Chương 2: Định nghĩa cũng như khái niệm về Mono-alphabetic Mono-alphabetic Substitution Cipher (Mật mã thay thế một chữ cái)

- Chương 3: Định nghĩa, khái niệm, ví dụ về Frequency Analysis (Phân tích tần số)

- Chương 4: Trình bày chứng mình code về mã hoá bằng Mono-alphabetic và giải mã bằng Frequency Analysis.

Những nội dụng câu hỏi đã qua chương trình học online và cung cấp cho sinh viên đầy đủ kiến thức để thực hiện bài làm một cách chỉnh chu nhất. Đa phần cũng cần phải tìm hiểu thêm một số kiến thức mở rộng khác trên các trang mạng. Sinh viên cần xem lại các lý thuyết về định nghĩa, định lí, các tính chất, các ví dụ, các bài tập, nghiên cứu các trang mạng truyền tải những thông tin liên quan đến môn học… từ đó ghi nhớ lại các kiến thức cần đạt để vận dụng vào bài báo cáo.

Qua bài báo cáo em thấy em nhận thấy đề tài vừa sức với mọi sinh viên theo học, và thông qua bài báo cáo em nhận thấy bản thân đã nổ lực tìm tòi, học hỏi nhiệt tình để có kiến thức hổ trợ cho quá trình thực hành, nhờ đó đã giúp cá nhân của mỗi sinh viên có thể khắc được các kiến thức về môn học kĩ hơn và sâu hơn. Cuối lời em xin gửi lời cảm ơn chân thành và sâu sắc nhất đến các thầy cô của khóa học đã giúp các bạn sinh viên có thể vượt qua bài báo cáo này với nhiều kĩ năng mới.

MỤC LỤC

[LỜI CẢM ƠN 2](#_Toc101397613)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN 4](#_Toc101397614)

[TÓM TẮT 5](#_Toc101397615)

[MỤC LỤC 6](#_Toc101397616)

[CHƯƠNG 1 – GIỚI THIỆU 8](#_Toc101397617)

[1. Mã hoá (encryption) 8](#_Toc101397618)

[2. Giải mã (decryption) 8](#_Toc101397619)

[3. Hệ thống mật mã đối xứng (symmetric cryptosystem) 9](#_Toc101397620)

[4. Hệ thống mật mã không đối xứng (asymmetric cryptosystem) 10](#_Toc101397621)

[CHƯƠNG 2 – MẬT MÃ THAY THẾ MỘT CHỮ CÁI (Mono-alphabetic Substitution Cipher) 11](#_Toc101397622)

[1. Định nghĩa 11](#_Toc101397623)

[2. Ví dụ 11](#_Toc101397624)

[3. Đánh giá 11](#_Toc101397625)

[CHƯƠNG 3 – PHÂN TÍCH TẦN SỐ (Frequency Analysis) 12](#_Toc101397626)

[1. Định nghĩa 12](#_Toc101397627)

[2. Ví dụ 13](#_Toc101397628)

[3. Đánh giá 15](#_Toc101397629)

[CHƯƠNG 4 – THỬ NGHIỆM 16](#_Toc101397630)

[1. Code mã hoá bằng mono-alphabetic (Encryption using Monoalphabetic Substitution Cipher) 16](#_Toc101397631)

[2. Code giải mã bằng phân tích tần số (Decryption using Frequency Analysis) 18](#_Toc101397632)

[TÀI LIỆU THAM KHẢO 21](#_Toc101397633)

CHƯƠNG 1 – GIỚI THIỆU

1. Mã hoá (encryption)

- Mã hóa là một phương tiện bảo mật dữ liệu kỹ thuật số bằng một hoặc nhiều kỹ thuật toán học, cùng với mật khẩu hoặc "khóa" được sử dụng để giải mã thông tin. Quá trình mã hóa dịch thông tin bằng cách sử dụng một thuật toán làm cho thông tin gốc không thể đọc được.

- Mã hóa là một cách quan trọng để các cá nhân và công ty bảo vệ thông tin nhạy cảm khỏi các cuộc tấn công mạng, hacker.

* Phân loại:

+ Mã hoá bất đối xứng

+ Mã hoá đối xứng

* Ví dụ:

- Các trang web truyền số thẻ tín dụng và tài khoản ngân hàng phải luôn mã hóa thông tin này để ngăn chặn hành vi trộm cắp danh tính và gian lận.

1. Giải mã (decryption)

**-** Giải mã là quá trình chuyển dữ liệu đã được mã hóa không thể đọc được trở lại dưới dạng mã hóa ban đầu của nó. Trong quá trình giải mã, hệ thống trích xuất và chuyển đổi dữ liệu bị cắt xén và chuyển nó thành văn bản và hình ảnh mà không chỉ người đọc mà còn cả hệ thống dễ hiểu. Việc giải mã có thể được thực hiện thủ công hoặc tự động.

- Giải mã là cần thiết vì:

+ Giúp bảo mật thông tin nhạy cảm như thông tin đăng nhập như tên người dùng và mật khẩu.

+ Cung cấp tính bảo mật cho dữ liệu cá nhân.

+ Giúp đảm bảo rằng bản ghi hoặc tệp không thay đổi.

+ Tránh đạo văn và bảo vệ quyền sở hữu trí tuệ.

+ Có lợi cho các giao tiếp mạng như internet, nơi tin tặc có thể truy cập vào dữ liệu không được mã hóa.

+ Cho phép một người bảo vệ dữ liệu của họ một cách an toàn mà không sợ người khác truy cập vào nó.

- Các loại giải mã:

+ Tripple DES

+ RSA

+ Blowfish

+ Twofish

+ AES

1. Hệ thống mật mã đối xứng (symmetric cryptosystem)

- Hệ thống mật mã đối xứng chỉ liên quan đến một khóa mà hệ thống sử dụng để mã hóa và giải mã. Người gửi và người nhận cần biết khóa trước khi trao đổi thông điệp. hệ thống sử dụng các giao thức trao đổi khóa (như Giao thức Diffie-Hellman) để đảm bảo rằng người gửi và người nhận đồng ý về khóa bí mật trước khi bắt đầu giao tiếp.

A picture containing diagram

Description automatically generated- Hạn chế của hệ thống này là người gửi và người nhận cần phải trao đổi khóa trước khi giải mã thông điệp. Nếu các khóa không được thay đổi thường xuyên, hệ thống sẽ dễ bị tấn công vì kẻ tấn công có thể sử dụng khóa bị rò rỉ để làm gián đoạn liên lạc.

1. Hệ thống mật mã không đối xứng (asymmetric cryptosystem)

- Hệ thống mật mã không đối xứng là một loại mã hóa sử dụng hai khóa riêng biệt nhưng có liên quan đến toán học để mã hóa và giải mã dữ liệu. Điểm mạnh của bảo mật nằm ở các thuộc tính của các khóa này vì nó không khả thi về mặt tính toán để tính toán một khóa này bằng cách sử dụng khóa kia. Mỗi người gửi và người nhận sẽ có cặp khóa riêng tư của họ trong hệ thống này.

Diagram

Description automatically generatedVí dụ: A muốn gửi một tin nhắn cho B, A sẽ cần sử dụng khóa công khai (public key) của B để mã hóa tin nhắn và B sẽ giải mã tin nhắn bằng khóa riêng (private key) của họ.

CHƯƠNG 2 – MẬT MÃ THAY THẾ MỘT CHỮ CÁI (Mono-alphabetic Substitution Cipher)

1. Định nghĩa

- Mật mã đơn chữ cái là một mật mã thay thế trong đó mỗi chữ cái của văn bản thuần túy được thay thế bằng một chữ cái khác của bảng chữ cái. Nó sử dụng một phím cố định bao gồm 26 chữ cái của một "bảng chữ cái ngẫu nhiên".

- Loại mật mã này là một dạng mã hóa đối xứng vì cùng một khóa có thể được sử dụng để vừa mã hóa vừa giải mã một thông điệp.

- Mật mã thay thế một chữ cái, nếu muốn giải mã được sau khi chúng bị mã hoá buộc người nhận phải biết được mã khoá của người gửi – mã khoá mà người bên đã thống nhất từ đầu, mới có thể giải mã được nội dung họ cần trao đổi.

1. Ví dụ

- Ta có mã khoá như sau:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Plain text alphabet: | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
| Cipher text alphabet (key): | E | I | J | V | P | G | N | C | A | S | R | T | H | X | K | O | F | M | W | Q | Y | U | L | B | Z | D |

- Plain text: Today, I start studying the first day at TDTU University, I am feeling so interested.

Sau khi dùng mã khoá trên và chuyển đổi ta được câu mã hoá sau:

🢥 QKVEZ, A WQEMQ WQYVZAXN QCP GAMWQ VEZ EQ QVQY YXAUPMWAQZ, A EH GPPTAXN WK AXQPMPWQPV

1. Đánh giá

- Đây là một mật mã thay thế đơn giản và tiện lợi trong nhiều trường hợp cho các thời chiến trước đây, tuy nhiên với tiêu chuẩn với sự phát triển của công nghệ ngày nay thì chúng rất yếu và cực kỳ dễ bị phá vỡ, nhưng chúng là một bước rất quan trọng trong việc phát triển mật mã.

CHƯƠNG 3 – PHÂN TÍCH TẦN SỐ (Frequency Analysis)

1. Định nghĩa

- Phân tích tần số là một trong những cách tấn công bản mã được biết đến . Phân tích tần số dựa trên việc nghiên cứu tần suất xuất hiện của các chữ cái hoặc nhóm chữ cái trong một bản mã.

- Ví dụ, trong tiếng Anh thường sử dụng các nguyên âm như e , o , a hoặc một phụ âm t . Mặt khác, có một số chữ cái rất hiếm, ví dụ như z hoặc x . Có các tuyên bố về tần số của các chữ cái trong các ngôn ngữ khác nhau . Các tần số chỉ có thể được xác định gần đúng bởi vì trong các loại văn bản khác nhau (khoa học, lịch sử, tiểu thuyết) chúng hơi khác nhau.

- Chẳng hạn như tần suất xuất hiện chữ cái trong tiếng anh như sau:

|  |  |  |  |
| --- | --- | --- | --- |
| a | 8.17% | n | 6.75% |
| b | 1.49% | o | 7.51% |
| c | 2.78% | p | 1.93% |
| d | 4.25% | q | 0.10% |
| e | 12.70% | r | 5.99% |
| f | 2.23% | s | 6.33% |
| g | 2.02% | t | 9.06% |
| h | 6.09% | u | 2.76% |
| i | 6.97% | v | 0.98% |
| j | 0.15% | w | 2.36% |
| k | 0.77% | x | 0.15% |
| l | 4.03% | y | 1.97% |
| m | 2.41% | z | 0.07% |

Chart

Description automatically generated

- Dựa vào đây ta có thể giải mã khoảng 30 – 40% nội dung mật mã cần giải và phải qua nhiều lần xem xét mới có thể rõ nội dung mật mã không thể giải được 100% chính xác nội dung ngay lần đầu được. Nội dung càng dài thì độ chính xác càng cao.

1. Ví dụ

Ta có mật mã sau: MIOL OL GFT GY MIT YTC COSR EKTAMWKTL MIAM BGW EAF LTT YKGD MIT MKAOF

🢥 Từ bảng mã trên ta đếm được tần suất xuất hiện các chữ cái như sau:

|  |  |  |
| --- | --- | --- |
| **T** | **8×** | **14.55%** |
| **M** | **7×** | **12.73%** |
| **K** | **4×** | **7.27%** |
| **O** | **4×** | **7.27%** |
| **L** | **4×** | **7.27%** |
| **G** | **4×** | **7.27%** |
| **I** | **4×** | **7.27%** |
| **A** | **4×** | **7.27%** |
| **F** | **3×** | **5.45%** |
| **Y** | **3×** | **5.45%** |
| **W** | **2×** | **3.64%** |
| **C** | **2×** | **3.64%** |
| **E** | **2×** | **3.64%** |
| **R** | **1×** | **1.82%** |
| **S** | **1×** | **1.82%** |
| **B** | **1×** | **1.82%** |
| **D** | **1×** | **1.82%** |

Với bảng mã tiếng anh: “ETAOINSHRDLCUMWFGYPBVKJXQZ” thì ta có mã khoá tương ứng nhau sau:

|  |  |
| --- | --- |
| **T** | **E** |
| **M** | **T** |
| **K** | **A** |
| **O** | **O** |
| **L** | **I** |
| **G** | **N** |
| **I** | **S** |
| **A** | **H** |
| **F** | **R** |
| **Y** | **D** |
| **W** | **L** |
| **C** | **C** |
| **E** | **U** |
| **R** | **M** |
| **S** | **W** |
| **B** | **F** |
| **D** | **G** |

- Tiến hành giải mã bằng cách thay kí tự ở mật mã thành kí tự ở mã khoá, ta thu được như sau:

MIOL OL GFT GY MIT YTC COSR EKTAMWKTL MIAM BGW EAF LTT YKGD MIT MKAOF

🢥 TSOI OI NRE ND TSE DEC COWM UAEHTLAEI TSHT FNL UHR IEE DANG TSE TAHOR

- Đây là kết quả sau khi giải mã

1. Đánh giá

- Với nội dung đoạn mã ngắn thì việc giải mã sẽ vô cùng khó khăn với phuong pháp ‘phân tích tần số’ này. Nhưng nếu đoạn mật mã dài thiệt sẽ giải rất dễ dàng để ra kết quả đúng nhất tuy nhiên sẽ mất rất nhiều thời gian.

CHƯƠNG 4 – THỬ NGHIỆM

1. Code mã hoá bằng mono-alphabetic (Encryption using Monoalphabetic Substitution Cipher)
2. def key():
3. key = 'EIJVPGNCASRTHXKOFMWQYULBZD'
4. fre = {}
5. for ascii in range(65,91):
6. fre[chr(ascii)] = key[ascii-65]
7. for ascii in range(97,122):
8. fre[chr(ascii)] = key[ascii-97]
9. return fre
10. print(key())
11. def encrypt\_monoalpha(message, monoalpha\_cipher):
12. encrypted\_message = []
13. for letter in message:
14. encrypted\_message.append(monoalpha\_cipher.get(letter, letter))
15. return ''.join(encrypted\_message)
16. message = "Today, I start studying the first day at TDTU University, I am feeling so interested"
17. print("Message: ",message)
18. print("After encrypt: ",encrypt\_monoalpha(message,key()))

Text

Description automatically generated🢥 Sau khi chạy code trên ta được:

- Ví dụ ở phần code trên được lấy trực tiếp từ phần lý thuyết phía trên ở Chương 2 để có thể dễ hình dung và chứng mình kết quả trên là đúng

- Fuction “**Key**” ở đây được dùng để ráng từng kí tự trong bảng chữ cái alpha bằng với bảng chữ cái mã khoá được đặt tên “key” – tức là mã khoá để có thể mã hoá được message theo mật mã Cipher text alphabet đã cho sẳn trước đó.

+ Dòng for thứ nhất tức là để ráng cái chữ cái in hoa trong bảng mã ascii bằng với chữ cái trong “key” vì ở bảng mã ascii từ 65 đến 91 có nghĩ là chữ cái in hoa từ ‘A’ đến ‘Z’.

+ Dòng for thứ nhất tức là để ráng cái chữ cái viết thường trong bảng mã ascii bằng với chữ cái trong “key” vì ở bảng mã ascii từ 97 đến 122 có nghĩ là chữ cái in viết thường từ ‘a’ đến ‘z’.

+ Sau khi ráng thành công ta đưa kết quả đã được ráng bằng nhau đó vào mảng tên ‘fre’ để dễ dàng dùng hoặc in ra xem. Kết quả của ‘fre’ đươc in ra ở phần kết quả phía trên.

- Function “**encrypt\_monoalpha(message, monoalpha\_cipher)**” được dùng để giải mã message từ nội dung ban đầu thành một message đã được mã hoá theo ‘key’ được tạo ra ở function trên.

+ Dòng for để kiểm tra từng kí tự trong message, khi gặp kí nào thuộc bảng chữ cái alpha sẽ thay đổi chúng thành kí tự trong mật mã mã hoá đã tạo trước đó. Sau khi thay đổi kí tự thành công sẽ đưa kí tự đó vào mảng ‘encrypted\_message’. Xét từng kí tự trong message cho đến khi hết chuỗi message đó.

+ Sau cùng thu được kết quả đã được mã hoá và lưu trong ‘encrypted\_message’ nhưng chúng ở khoảng cách xa nhau vì ở phần for xét từng kí tự mới lưu vào mảng, nên ta dùng lên join đến đưa chung lại gần nhau.

+ Kết quả sau cùng được in ra ở phần sau khi chạy code phía trên.

1. Code giải mã bằng phân tích tần số (Decryption using Frequency Analysis)
2. #Decryption using Frequency Analysis.
3. def number (frequency\_analysis):
4. return frequency\_analysis[1]
6. def decrypt\_frequencyanalysis(massage):
7. #Tần suất xuất hiện các kí tự trong mật mã bị mã hoá
8. frequency\_analysis = { "A" : 0,  "B" : 0,  "C" : 0,  "D" : 0,  "E" : 0,"F" : 0,  "G" : 0,
9. "H" : 0,  "I" : 0,  "J" : 0,  "K" : 0,  "L" : 0,  "M" : 0,  "N" : 0,  "O" :   0,
10. "P" : 0,  "Q" : 0,  "R" : 0,  "S" : 0,  "T" : 0,  "U" : 0,  "V" : 0,  "W" : 0,
11. "X" : 0,  "Y" : 0,  "Z" : 0 }
12. for letter in massage:
13. if letter.isalpha():
14. frequency\_analysis[letter] += 1
15. #Sắp xếp lại tần suất theo thứ tự giảm dần
16. unsorted = frequency\_analysis.items()
17. descend = list(sorted(unsorted, key = number, reverse=True))
18. #Engsub tượng trưng cho tân suất xuất hiện của bảng chữ cái ở tiếng Anh
19. engsub = "ETAOINSHRDLCUMWFGYPBVKJXQZ"
20. #Ráng kí tự có tần suất xuất hiện cao nhất trong mật mã theo thứ tự của engsub
21. fre = {}
22. for i in range(0, 26):
23. fre[descend[i][0]] = engsub[i]
24. #Tiến hành thay thế kí tự đã ráng phía trên vào mật mã thu được mật mã giải mã lần thứ 1
25. decryptedlist = []
26. for letter in massage:
27. asciitext = ord(letter.upper())
28. if  asciitext <= 90 and asciitext >= 65:
29. decryptedlist.append(fre[letter])
30. decrypted\_message = "".join(decryptedlist)
31. return decrypted\_message
32. #50 words
33. message1 = "Whenever travellers penetrate into remote regions where human hunters are unknown, they find the wild things half tame, little afraid of man, and inclined to stare curiously from a distance of a few paces. But very soon they learn that man is their most dangerous enemy, and fly from him as soon as he is seen. It takes a long time and much restraint to win back their confidence"
34. encrypted\_text1 = encrypt\_monoalpha(message1,key())
35. print("\*\*\*encrypt1 - 50 words: ", encrypt\_monoalpha(message1,key()))
36. print("\_\_\_decrypt1:",decrypt\_frequencyanalysis(encrypted\_text1))
37. print()

Text

Description automatically generated🢥 Code sau khi chạy với mật mã 50 từ như message1:

- Function ‘**decrypt\_frequencyanalysis(text):**’ Dùng để giải mã mật mã đã bị mã hoá bằng phương thức frequency analysis (phân tích tần số).

+ Mảng **frequency\_analysis** chứa chữ cái với tần suất xuất hiện của chúng ban đầu khi chưa có mật mã nào, thì mỗi chữ cái đều có tần suất là 0

+ Dòng for số 32 dùng để đếm từng kí tự trong mật mã mã hoá. Khi có chữ cái nào là alpha(trong bảng chữ cái mặc định) xuất hiện thì sẽ tăng số tần suất xuất hiện của chúng trong mảng **frequency\_analysis** lên 1 đơn vị. Cứ xét như thế cho đến hết mật mã truyền vào.

+ Dòng thứ 38, unsorted là mảng lưu lại các giá trị trong **frequency\_analysis** với một định dạng khác

+Dòng thứ 39, **descend** sẽ sắp xếp lại thứ tự xuất hiện của các chữ cái theo tần suất xuất hiện của mỗi chữ cái theo thứ tự giảm dần dựa trên giá trị của tần suất của mỗi chữ cái ở **Unsorted**

**+** Function “**Number”** dùng để lấy giá trị của item (chữ cái) phía sau dấu “:” ở mảng **Unsorted**

+ Engsub là mảng chứ tần suất của chữ cái theo thứ tự giảm dần ở **Tiếng Anh.**

+ Dòng for thứ 45, xét 26 lần tương đương 26 chữ cái trong bảng chữ cái. Theo thứ tự tăng dần, chữ cái đầu tiên của mảng **Descend** sẽ bằng với chữ cái đầu tiên trong mảng **Engsub**. Xếp các giá trị đã được ráng đó vào mảng tên “**fre**”

+ Dòng for thứ 49, xét từng kí tứ của mật mã đã mã hoá. Xét đến chữ cái nào dùng lệnh ‘ord’ để đưa kí tự đó về dạng số ở bảng mã ascii. Nếu số thuộc trong khoảng [65, 90] tức là nằm trong khoảng là chữ cái in hoa ở bảng mã ascii thì đưa kí tự đó vào mảng tên **decryptedlist.** Xét cho đến khi hết mật mã.

+ Vì khi được thêm vào mảng các kí tự ở một khoảng cách rất xa nhau, nên ta dùng **"".join(decryptedlist)** để đưa các kí tự lại gần với nhau và đưa chúng vào mảng tên **decrypted\_message.**

**+** Kết quả cuối cùng thu được với ví dụ 50 từ mã hoá được in ra ở phần kết quả phía trên.

- Các ví dụ khác như 100 từ, 500 từ, 1000 từ,…v.v. Đã được ghi rõ trong phần code hơn vì nội dung đó khi copy qua word sẽ rất dài, mong thầy có thể xem rõ hơn ở phần code.

TÀI LIỆU THAM KHẢO

**Tiếng Việt**

1. Giới thiệu về mật mã cổ điển, 2014, WhiteHat, access <https://whitehat.vn/threads/gioi-thieu-ve-mat-ma-co-dien-phan-3-mat-ma-thay-the.2970/>
2. Markhermy3100 , 2021, “Cryptography Cổ Điển - Mã Thay Thế*” access:*[*https://codelearn.io/sharing/cryptography-co-dien-ma-thay-the-2*](https://codelearn.io/sharing/cryptography-co-dien-ma-thay-the-2)
3. Lê Tôn Phát, 2020, Những khái niệm cơ bản trong Mật mã học, access <https://cafedev.vn/ctdl-stack-vi-du-ve-hoan-doi-toan-tu-trong-mot-phep-tinhinfix-to-postfix/>

……….

**Tiếng Anh**

1. Wikipedia, Frequency analysis , 2021, access <https://en.wikipedia.org/wiki/Frequency_analysis>
2. CRYPTO CORNER, Monoalphabetic Substitution Ciphers, access<https://crypto.interactive-maths.com/monoalphabetic-substitution-ciphers.html>
3. Bruce Ikenaga, 2019, Traditional Ciphers, access <https://www.tutorialspoint.com/cryptography/traditional_ciphers.htm#:~:text=Monoalphabetic%20cipher%20is%20a%20substitution,get%20encrypted%20to%20'D'.>
4. Frequency Analysis in Python, 2018, access <https://www.codedrome.com/frequency-analysis-in-python/>
5. Momo, Caesar Cipher And Frequency Analysis With Python, 2021, access <https://medium.com/@momohakarish/caesar-cipher-and-frequency-analysis-with-python-635b04e0186f>

………..