TRƯỜNG ĐH SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN

Logo

Description automatically generated

**TIỂU LUẬN CUỐI KỲ**

**Môn học: Cấu trúc dữ liệu và giải thuật**

**Xây dựng trò chơi nối từ Tiếng Việt**

Giảng viên: PGS.TS. Hoàng Văn Dũng

**Danh sách sinh viên thực hiện**

|  |  |  |
| --- | --- | --- |
| **Mã số SV** | **Họ và tên** | **Tỷ lệ**  **đóng góp** |
| 20133079 | Lã Minh Phúc | 120% |
|  |  |  |
|  |  |  |

*TP. Hồ Chí Minh, tháng .... năm .......*

# MỤC LỤC

# MỞ ĐẦU

Ngày nay với sự phát triển của internet thì mạng xã hội đã trở thành một phần trong cuộc sống mỗi chúng ta. Sử dụng mạng xã hội trong mục các mục đích như giao tiếp, chia sẻ, giải trí giúp chúng ta luôn cập nhập thông tin một cách nhanh chóng. Tuy nhiên chính vì nhu cầu trao đổi thông tin nhanh chóng nên nhiều người dùng đã sử dụng các từ vay mượn từ tiếng nước ngoài trong quá trình chia sẻ thông tin nhằm rút ngắn thời gian nhập liệu. Điều này dẫn tới việc lạm dụng từ mượn làm mất đi vẻ đẹp và sự phong phú của Tiếng Việt khi ngày càng nhiều các từ Tiếng Việt dần bị thay thế như “tạm biệt”, “xin chào”,

# CƠ SỞ LÝ THUYẾT

## Công cụ và môi trường để lập trình

### Xây dựng chương trình

* Visual Studio Community 2019, Desktop Development with C++

### Xử lý, phân tích dữ liệu

* Python 3.8
* Anaconda, Jupyter Notebook
* Từ điển Việt Việt; Hồ Ngọc Đức

### Trao đổi, thảo luận

* Git, Github
* Facebook, Messenger

## Các phương pháp, kĩ thuật sử dụng

* Danh sách sử dụng mảng
* Danh sách sử dụng liên kết đơn
* Bảng băm
* Tìm kiếm nhị phân

# PHÂN TÍCH, THIẾT KẾ GIẢI PHÁP

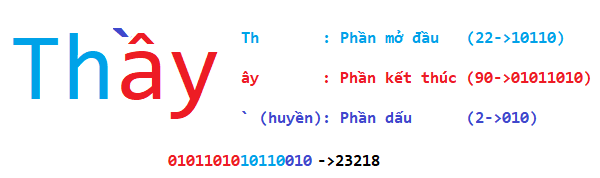
## Thuật toán mã một âm trong Tiếng Việt

Phân tách một âm trong Tiếng Việt thành 3 thành phần: phần bắt đầu, phần kết thúc và phần dấu.

Sử dụng dữ liệu unsigned short (uint16\_t) để lưu trữ 3 thành phần này, cụ thể:

* 8 bits đầu: Phần kết thúc
* 5 bits tiếp theo: Phần bắt đầu
* 3 bits còn lại: Phần dấu

Các âm đặc biệt như “gi, yên, yêm, yêu” được tính phần bắt đầu là 31, phần kết thúc là giá trị mã hoá của âm.



Hình Nal: Mô tả phân tách và mã hoá âm “thầy”

# TÀI LIỆU THAM KHẢO

* Hồ Ngọc Đức (2004), Từ điển tiếng Việt, <http://informatik.uni-leipzig.de/~duc/Dict/>. Dữ liệu từ điển của trò chơi

# PHỤ LỤC

## Giá trị mã hoá các thành phần của âm

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Phần bắt đầu** | **Giá trị** | **Phần bắt đầu** | **Giá trị** | **Phần bắt đầu** | **Giá trị** |
|  | 0 | k | 9 | q | 18 |
| b | 1 | kh | 10 | r | 19 |
| c | 2 | l | 11 | s | 20 |
| ch | 3 | m | 12 | t | 21 |
| d | 4 | n | 13 | th | 22 |
| g | 5 | ng | 14 | tr | 23 |
| gh | 6 | ngh | 15 | v | 24 |
| gi | 7 | nh | 16 | x | 25 |
| h | 8 | ph | 17 | đ | 26 |

Bảng Nal: Giá trị mã hoá phần bắt đầu

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Phần kết thúc** | **Giá trị** | **Phần kết thúc** | **Giá trị** | **Phần kết thúc** | **Giá trị** |
| a | 0 | oc | 45 | ây | 90 |
| ac | 1 | oe | 46 | ê | 91 |
| ach | 2 | oi | 47 | êch | 92 |
| ai | 3 | om | 48 | êm | 93 |
| am | 4 | on | 49 | ên | 94 |
| an | 5 | ong | 50 | ênh | 95 |
| ang | 6 | op | 51 | êp | 96 |
| anh | 7 | ot | 52 | êt | 97 |
| ao | 8 | oăn | 53 | êu | 98 |
| ap | 9 | oăng | 54 | ô | 99 |
| at | 10 | oăt | 55 | ôc | 100 |
| au | 11 | u | 56 | ôi | 101 |
| ay | 12 | ua | 57 | ôm | 102 |
| e | 13 | uc | 58 | ôn | 103 |
| em | 14 | ui | 59 | ông | 104 |
| en | 15 | um | 60 | ôp | 105 |
| eng | 16 | un | 61 | ôt | 106 |
| eo | 17 | ung | 62 | ăc | 107 |
| ep | 18 | up | 63 | ăm | 108 |
| et | 19 | ut | 64 | ăn | 109 |
| i | 20 | uy | 65 | ăng | 110 |
| ia | 21 | uya | 66 | ăp | 111 |
| ich | 22 | uych | 67 | ăt | 112 |
| im | 23 | uynh | 68 | ơ | 113 |
| in | 24 | uyt | 69 | ơi | 114 |
| inh | 25 | uyên | 70 | ơm | 115 |
| ip | 26 | uyêt | 71 | ơn | 116 |
| it | 27 | uân | 72 | ơp | 117 |
| iu | 28 | uât | 73 | ơt | 118 |
| iêc | 29 | uê | 74 | ư | 119 |
| iêm | 30 | uôc | 75 | ưa | 120 |
| iên | 31 | uôi | 76 | ưc | 121 |
| iêng | 32 | uôm | 77 | ưi | 122 |
| iêp | 33 | uôn | 78 | ưng | 123 |
| iêt | 34 | uông | 79 | ưt | 124 |
| iêu | 35 | uôt | 80 | ưu | 125 |
| o | 36 | uơ | 81 | ươc | 126 |
| oa | 37 | y | 82 | ươi | 127 |
| oach | 38 | âc | 83 | ươm | 128 |
| oai | 39 | âm | 84 | ươn | 129 |
| oan | 40 | ân | 85 | ương | 130 |
| oang | 41 | âng | 86 | ươp | 131 |
| oanh | 42 | âp | 87 | ươt | 132 |
| oat | 43 | ât | 88 | ươu | 133 |
| oay | 44 | âu | 89 |  |  |

Bảng Nal: Giá trị mã hoá phần kết thúc

|  |  |
| --- | --- |
| **Phần dấu** | **Giá trị** |
| Thanh | 0 |
| Sắc | 1 |
| Huyền | 2 |
| Hỏi | 3 |
| Ngã | 4 |
| Nặng | 5 |

Bảng Nal: Giá trị mã hoá phần dấu

|  |  |
| --- | --- |
| **Âm đặc biệt** | **Giá trị** |
| gi | 0 |
| yên | 1 |
| yêm | 2 |
| yêu | 3 |

Bảng Nal: Giá trị mã hoá âm đặc biệt