

**ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**UIT**

# **BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**MÔN: NGÔN NGỮ LẬP TRÌNH C#**

## **PHẦN MỀM DỊCH THUẬT DTTRANSLATE**

**Sinh viên thực hiện:**

**Hoàng Tiến Dũng - 19521388**

**Nguyễn Thành Trọng - 19522410**

**Giảng viên: ThS. Nguyễn Bích Vân**

**Lớp : CS511.M11**

**NĂM 2021**

# MỤC LỤC

|   |    |
|---|----|
| <b>MỞ ĐẦU</b>   | 3  |
| <b>I. Giới thiệu phần mềm</b>                                       | 4  |
| <b>II. Mô tả chức năng của phần mềm</b>                             | 5  |
| 1. Dịch cơ bản  | 5  |
| 2. Dịch nâng cao  | 6  |
| 3. Lưu lịch sử tìm kiếm   | 6  |
| 4. Lưu từ quan trọng  | 7  |
| 5. Tìm kiếm và dịch các từ có trong hình ảnh                        | 8  |
| 6. Phát âm  | 8  |
| 7. Cài đặt  | 9  |
| <b>III. Mô tả dữ liệu</b>   | 9  |
| 1. Dữ liệu từ điển  | 9  |
| 2. Dữ liệu âm thanh   | 10 |
| 3. Dữ liệu Tesseract  | 10 |
| <b>IV. Hướng dẫn sử dụng phần mềm</b>                               | 10 |
| 1. Translation/Dịch   | 11 |
| 2. Advanced translation/Dịch nâng cao                               | 12 |
| 3. Search history/Lịch sử tìm kiếm                                  | 13 |
| 4. Important words/Từ quan trọng                                    | 14 |
| 5. OCR and Translate/OCR và dịch                                    | 15 |
| 6. Setting  | 16 |
| <b>V. Hướng phát triển</b>  | 17 |
| <b>VI. Mô tả công việc của các thành viên trong nhóm và đề xuất</b> | 17 |
| 1. Công việc của các thành viên                                     | 17 |
| 2. Đề xuất  | 18 |
| <b>VII. Nguồn học và tham khảo</b>                                  | 18 |

# MỞ ĐẦU

Hiện nay, tiếng Anh là được coi là ngôn ngữ quốc tế số một thế giới, nhu cầu đọc, hiểu tiếng Anh cũng rất cần thiết. Đặc biệt, đối với sinh viên, việc đọc và hiểu tiếng Anh có thể nói gần như là bắt buộc vì nó mở ra cho sinh viên vô số kho tàng kiến thức mới mẻ, nơi mà các bạn được hòa chung với nền kiến thức thế giới.

Với tầm quan trọng đó, nhóm chúng em gồm Hoàng Tiến Dũng và Nguyễn Thành Trọng đã quyết định làm đồ án với đề tài **Phần mềm dịch thuật DTTranslate**. Đây là phần mềm giúp giải quyết những vấn đề cơ bản về dịch thuật Việt – Anh, Anh - Việt. Bên cạnh đó, phần mềm còn có những chức năng giúp người dùng sử dụng phần mềm nhanh chóng và tiết kiệm thời gian nhất.

Bài viết dưới đây nhóm sẽ giới thiệu và phân tích **Phần mềm dịch thuật DTTranslate**. Tuy nhóm đã rất cố gắng cải thiện bài viết và phần mềm nhưng chắc chắn không thể tránh khỏi sự thiếu sót và sai lầm trong đó. Mong cô và những người đọc góp ý nếu thấy sai sót hoặc cần bổ sung.

Chúng em xin chân thành cảm ơn!

## I. Giới thiệu phần mềm

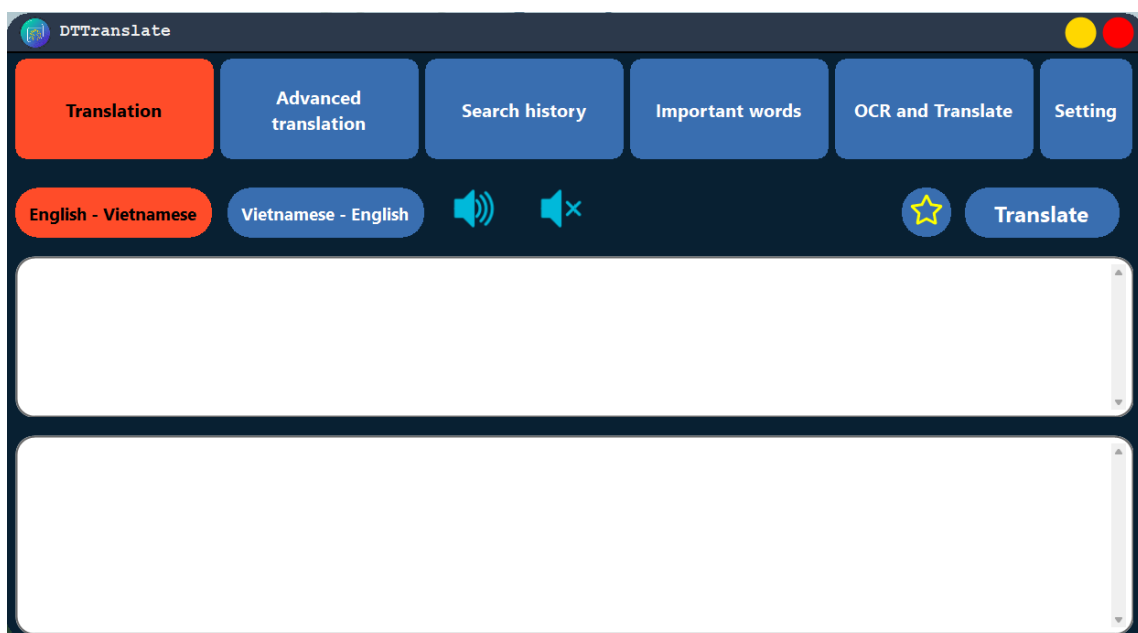
**DTTranslate** là phần mềm được xây dựng từ mong muốn được giúp mọi người Việt Nam có thể đọc, tìm hiểu tài liệu tiếng Anh và cũng như người nước ngoài muốn đọc, tìm hiểu tài liệu của Việt Nam.

Nhóm đã chọn Logo cho **DTTranslate** như ở dưới. Logo gồm biểu tượng Trái đất nằm trên biểu tượng Tài liệu. Điều đó muốn nói rằng, nguồn tri thức nhân loại, tất cả đều ở đó, nó là tài sản chung của nhân loại, sự khác biệt về ngôn ngữ không thể ngăn cản chúng ta tìm hiểu về nó. Màu xanh dương (nền) là màu của biển cả, có ý nghĩa rộng lớn có ý nghĩa là nguồn tri thức rộng lớn như đại dương cần chúng ta khám phá.



*Hình 1: Logo của DTTranslate*

Về giao diện bắt đầu, nhóm đã chọn cách hiển thị đơn giản, các nút bấm trực quan, dễ hiểu, rõ ràng với mục đích dễ tìm hiểu, tiếp cận cho người mới và dễ tiếp cận với nhiều tầng lớp (người trẻ, người lớn tuổi,...)



Hình 2: Giao diện khi khởi động DTTranslate

## II. Mô tả chức năng của phần mềm

Phần mềm DTTranslate gồm các chức năng nổi bật như:

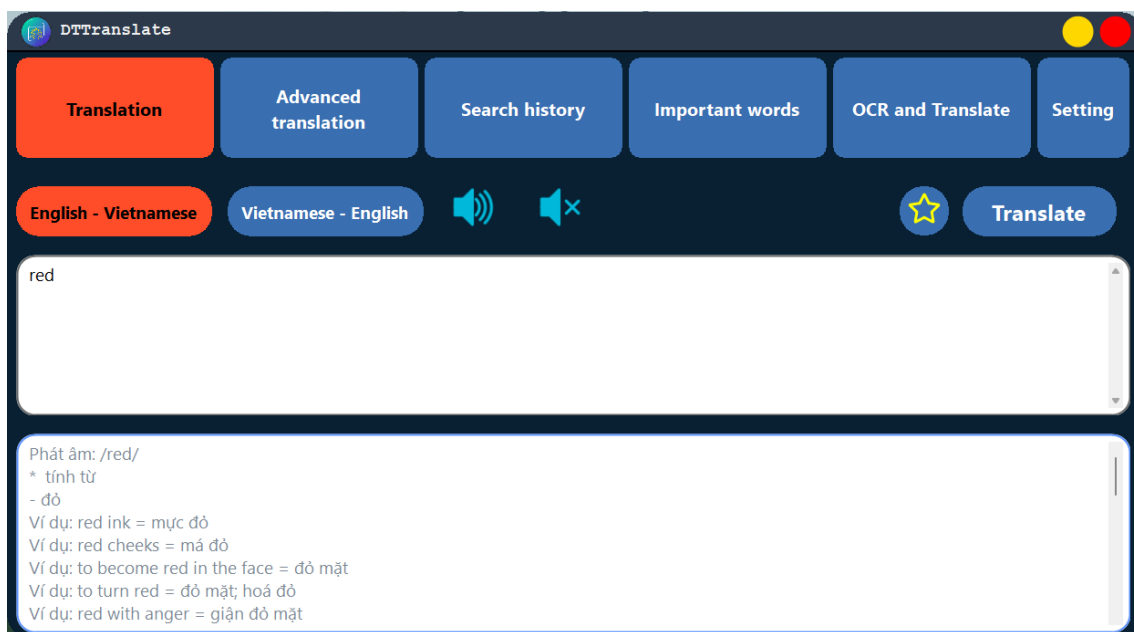
- Dịch cơ bản
- Dịch nâng cao
- Lưu lịch sử tìm kiếm
- Lưu từ quan trọng
- Tìm kiếm và dịch các từ có trong hình ảnh
- Phát âm
- Ngoài ra, phần mềm còn cho phép người dùng lựa chọn màu sắc phù hợp của giao diện, chọn ngôn ngữ hiển thị và chọn giọng đọc văn bản.

Sau đây sẽ là các mô tả chức năng nổi bật của phần mềm.

### 1. Dịch cơ bản

Phần mềm cho phép người dùng dịch từ tiếng Anh sang tiếng Việt và từ tiếng Việt sang tiếng Anh.

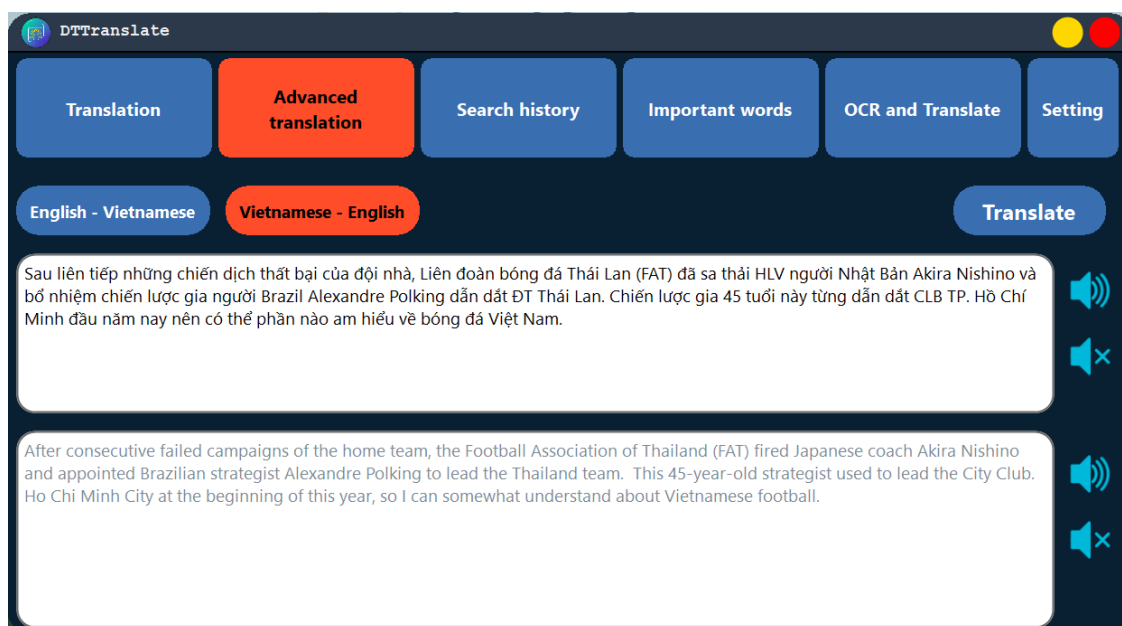
Một cách gọi khác của chức năng dịch cơ bản là dịch theo từ điển. Ở chức năng này, phần mềm sẽ cho biết các nghĩa và loại từ của từ muốn dịch, cách phát âm (Anh – Việt), các ví dụ và cụm từ thường gặp. Chức năng này như một từ điển mini, dùng khi người dùng muốn hiểu rõ nghĩa của từ đó.



*Hình 3: Mô tả chức năng Dịch cơ bản ở Anh – Việt*

## 2. Dịch nâng cao

Chức năng dịch nâng cao cho phép người dùng dịch một đoạn văn dài. Người dùng có thể dùng chức năng này để dịch một đoạn trong tờ báo hay dùng để dịch tài liệu Anh sang Việt hoặc Việt sang Anh.



*Hình 4: Mô tả chức năng Dịch nâng cao ở Việt - Anh*

## 3. Lưu lịch sử tìm kiếm

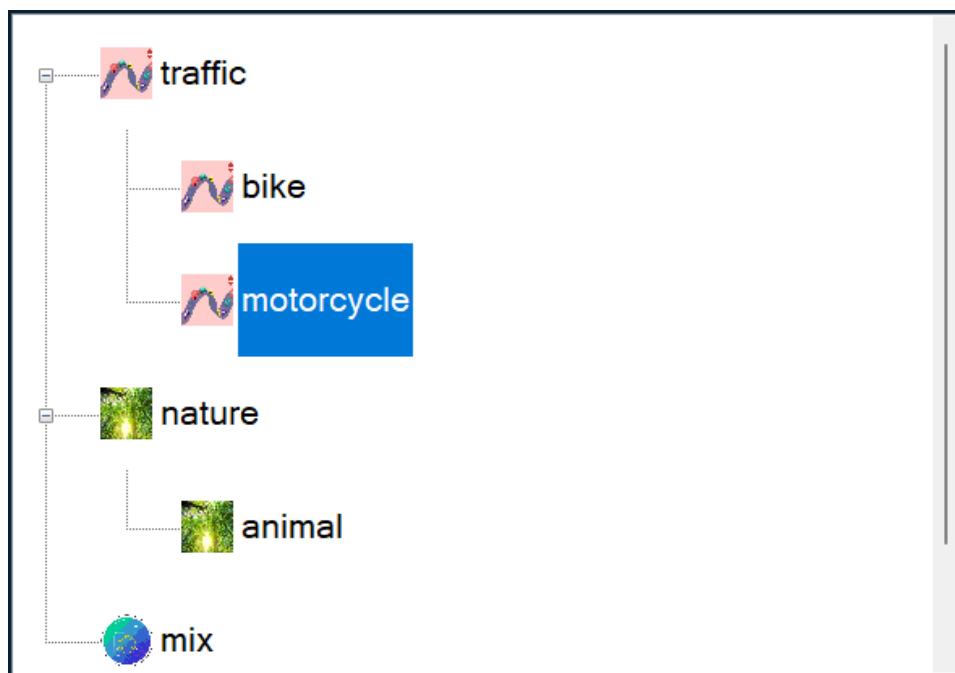
Lưu trữ lại lịch sử dịch (cơ bản) của người dùng giúp thuận tiện khi người dùng muốn nhớ lại những từ mình đã từng tra cứu.



Hình 5: Lưu lịch sử tìm kiếm

#### 4. Lưu từ quan trọng

Chức năng này đóng vai trò như một quyển sổ ghi chép, giúp người dùng tạo các danh mục và lưu các từ quan trọng vào đó. Lưu từ không những giúp người dùng lưu trữ mà còn giúp người dùng dễ nắm bắt và kiểm soát các từ trong việt ôn tập lại. Bên cạnh đó phần mềm còn cho chọn ảnh cho các mục làm các mục trực quan hơn.



Hình 6: Cấu trúc lưu trữ của chức năng Lưu từ quan trọng

## 5. Tìm kiếm và dịch các từ có trong hình ảnh

Chức năng cho phép người dùng chọn hình ảnh từ máy tính hoặc chụp một phần hình ảnh từ màn hình của người dùng. Từ hình ảnh đó, phần mềm sẽ phát hiện các đoạn chữ có trong ảnh và dịch.

Nếu bạn quan tâm đến vấn đề về tin tức thế giới, kinh doanh, chính trị Mỹ hay các bài viết về khoa học công nghệ thì nên tìm đến Newsweek. Đây là nguồn tài liệu vô cùng phong phú dành cho những bạn có mức tiếng Anh trung bình.



Hình 7: Ảnh đầu vào và kết quả

## 6. Phát âm



Chức năng Phát âm chuyển các chữ, đoạn chữ, hay văn bản thành một bài nói. Chức năng Phát âm kết hợp với chức năng Dịch nâng cao sẽ giúp người dùng nghe một bài viết nào đó hay thậm chí là một bài báo trước hoặc sau khi dịch sang ngôn ngữ mong muốn.

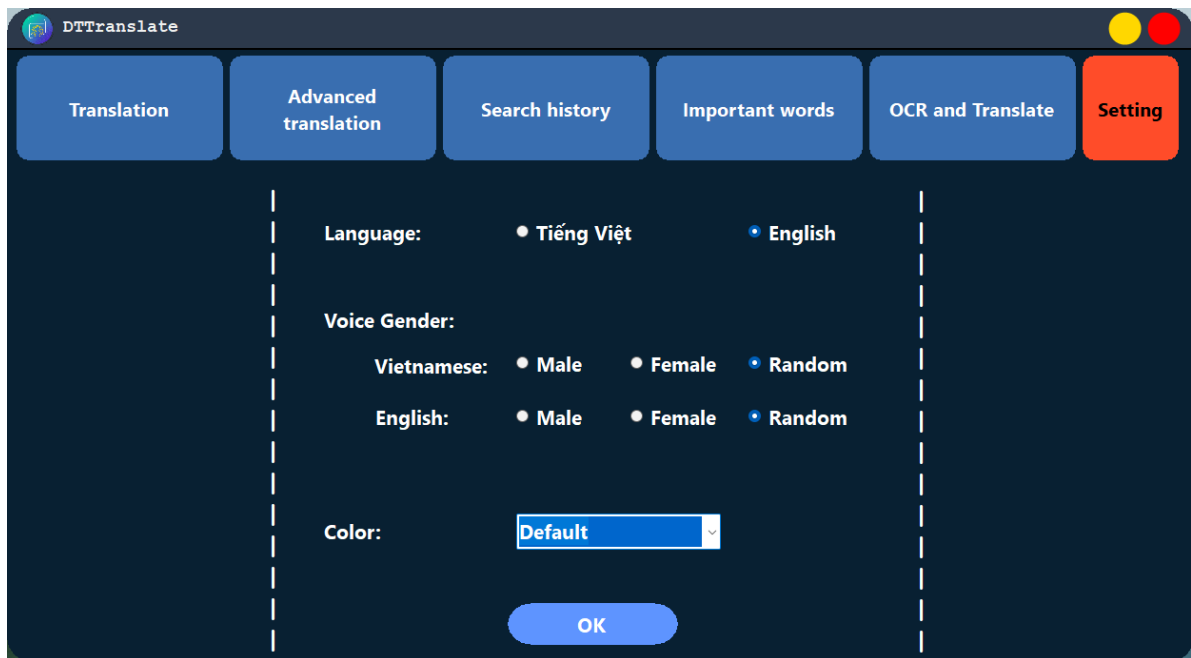


Hình 8: Nút bấm phát âm và dừng phát âm

## 7. Cài đặt

Ở phần cài đặt, phần mềm cho phép người dùng lựa chọn:

- Ngôn ngữ hiển thị, ngôn ngữ hiển thị trên các nút bấm,... là tiếng Việt hay tiếng Anh
- Giọng đọc cho từng tiếng gồm 3 loại: Nam, Nữ, Ngẫu nhiên
- Màu sắc cho giao diện, có 4 cặp màu sắc cho người dùng lựa chọn.



Hình 9: Giao diện phần cài đặt của phần mềm

## III. Mô tả dữ liệu

### 1. Dữ liệu từ điển

Về từ điển gồm từ điển Việt – Anh và từ điển Anh – Việt, nhóm đã tham khảo và thu thập dữ liệu từ điển từ hai nguồn là dự án từ điển của Hồ Ngọc Đức và dữ

liệu từ điển của Stardict.

Sau khi thu thập, nhóm đã tiến hành gộp và chuyển nó về một kiểu định dạng rồi tiến hành viết code bóc tách dữ liệu dạng text (từ, phát âm, ví dụ, cụm từ, thành ngữ,...) để đưa vào SQLite. Kết quả sau khi xử lý thu được là:

- Từ điển Anh – Việt: 58039 từ
- Từ điển Việt – Anh: 23495 từ

## **2. Dữ liệu âm thanh**

Ở phát âm tiếng Anh, nhóm đã sử dụng thư viện System.speech vì tốc độ xử lý nhanh và giọng đọc chuẩn phổ thông.

Ở phát âm tiếng Việt, nhóm đã sử dụng Viettel Text To Speech API. Sau khi gọi API bằng giao thức POST, trong file bin/debug sẽ xuất hiện một file wav. Đây chính là file âm thanh của đoạn text được gửi đi.

## **3. Dữ liệu Tesseract**

Nhóm sử dụng thư viện Tesseract cho phần OCR, với ảnh đầu vào là ảnh được lấy từ máy tính hoặc ảnh được chụp từ một phần của màn hình đã được chuyển xám và xóa nhiễu.

Để nhận diện được chữ, ta cần có hai file dữ liệu là vie.traineddata và eng.traineddata đã được huấn luyện trước.

# **IV. Hướng dẫn sử dụng phần mềm**

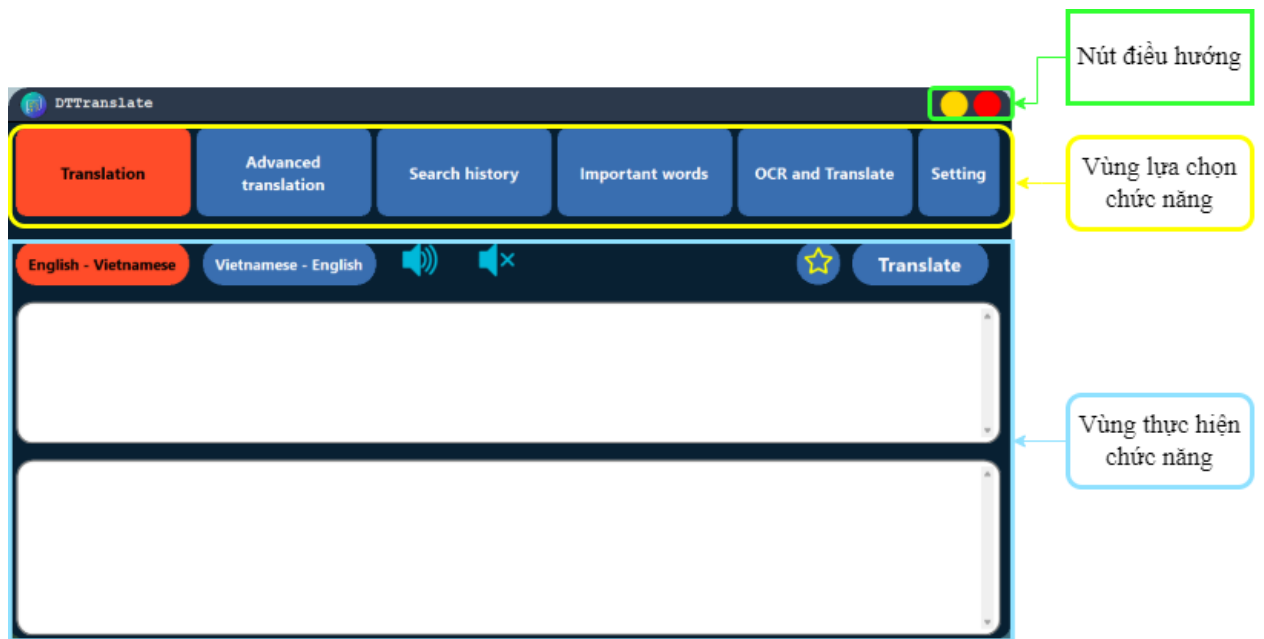
Như đã thấy ở trên giao diện khởi động mặc định của phần mềm đã ở phần Translation/Dịch. Phần mềm được chia thành 3 phần chính:

- Phần chứa nút điều hướng: nút đỏ - tắt chương trình, nút vàng - ẩn chương trình xuống thanh Taskbar.
- Phần chức năng: người dùng chọn chức năng mình muốn sử dụng.
- Phần thực hiện chức năng: là vùng người dùng thao tác trên đó để sử dụng chức năng.

Nhóm đã làm một video demo sử dụng phần mềm DTTranslate:

<https://www.youtube.com/watch?v=K0cUEGDI--s>

Dưới đây là cách dùng của DTTranslate được mô tả bằng hình ảnh.



Hình 10: Giao diện khởi động của phần mềm DTTranslate

## 1. Translation/Dịch

Phần Translation/Dịch là phần thực hiện chức năng Dịch cơ bản và Phát âm.

Cách sử dụng:

- Bước 1: Chọn button Translation/Dịch
- Bước 2: Chọn loại dịch muốn sử dụng: Anh – Việt, Việt – Anh
- Bước 3: Nhập từ muốn dịch vào phần text bên trên
- Bước 4: Chọn Translate/Dịch để bắt đầu dịch, kết quả sẽ được hiển thị phần text bên dưới.

Ngoài ra, người dùng click vào biểu tượng Loa để nghe phát âm từ được dịch và click vào biểu tượng ngôi sao để chuyển sang phần Lưu từ quan trọng nếu muốn lưu từ đó.



Hình 11: Cách sử dụng Translation/Dịch

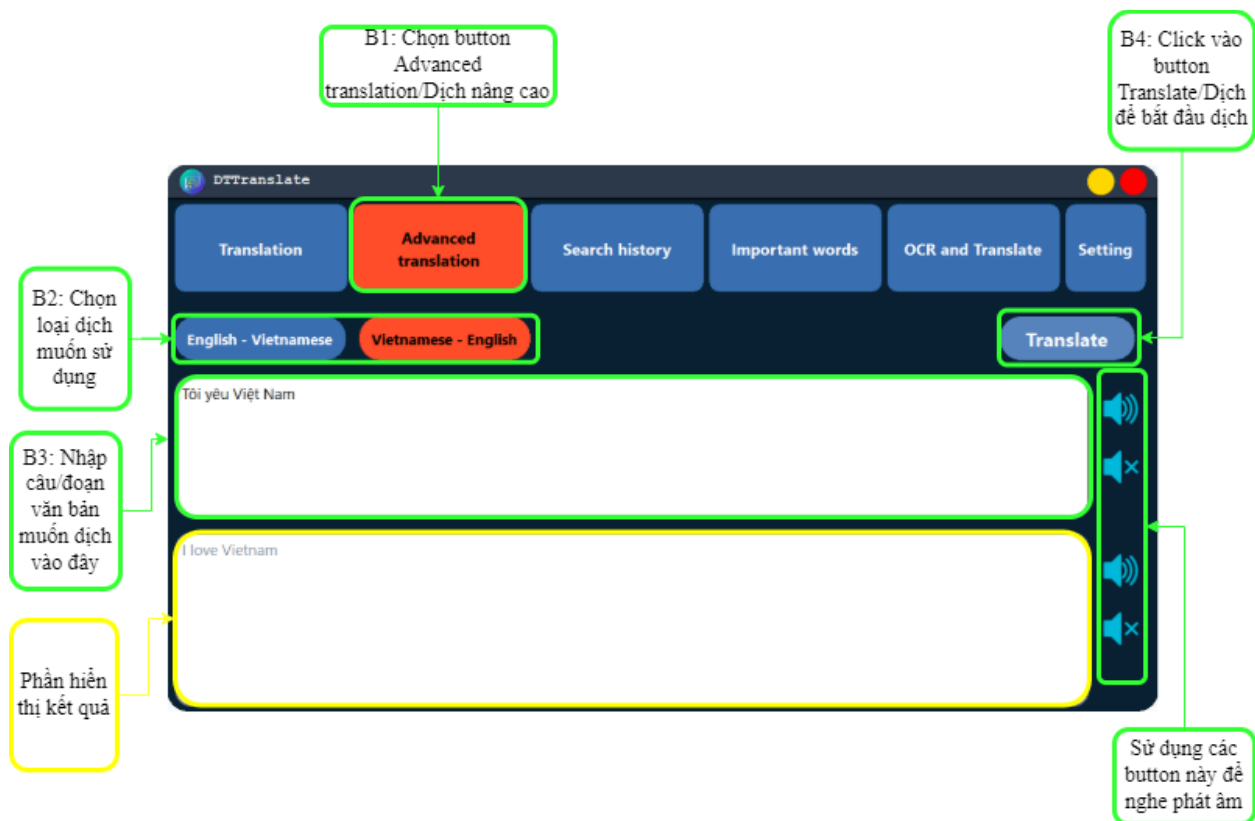
## 2. Advanced translation/Dịch nâng cao

Ở phần này, phần mềm thực hiện chức năng Dịch câu và Phát âm. Như đã nói ở trên, ở phần này có thể dịch được một câu dài hoặc một đoạn văn bản, thậm chí có thể cho máy đọc một bài báo trước hoặc sau khi dịch.

Cách sử dụng:

- Bước 1: Chọn button Advanced translation/Dịch nâng cao
- Bước 2: Chọn loại dịch muốn sử dụng: Anh – Việt, Việt – Anh
- Bước 3: Bước 3: Nhập từ muốn dịch vào phần text bên trên
- Bước 4: Chọn Translate/Dịch để bắt đầu dịch, kết quả sẽ được hiển thị phần text bên dưới.

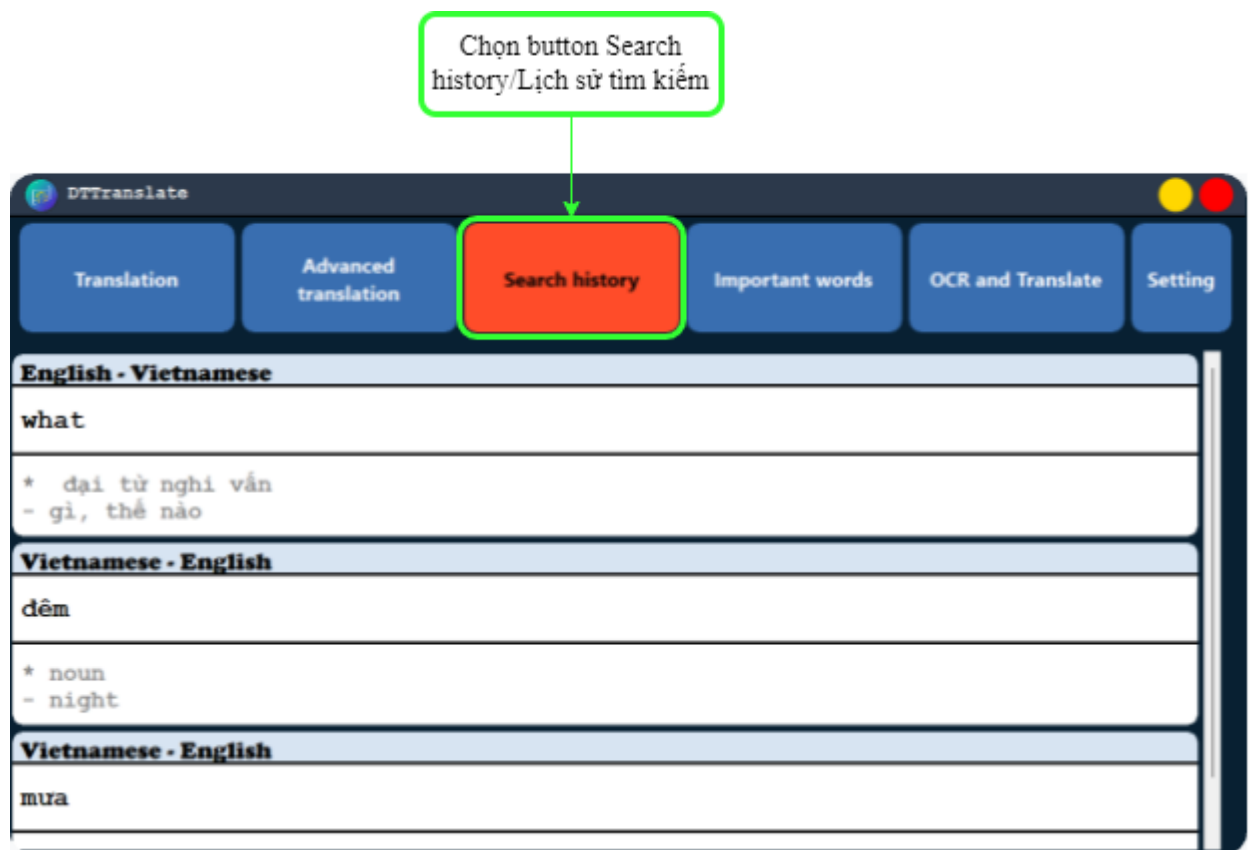
Ngoài ra, người dùng sử dụng các nút bấm biểu tượng hình Loa để nghe phát âm ở phần mong muốn.



Hình 12: Cách sử dụng Advanced translation/Dịch nâng cao

### 3. Search history/Lịch sử tìm kiếm

Phần lịch sử tìm kiếm là phần lưu những từ đã được tìm kiếm ở phần Translation/Dịch. Chỉ cần click vào Search history/Lịch sử tìm kiếm ta có thể chuyển sang phần này.



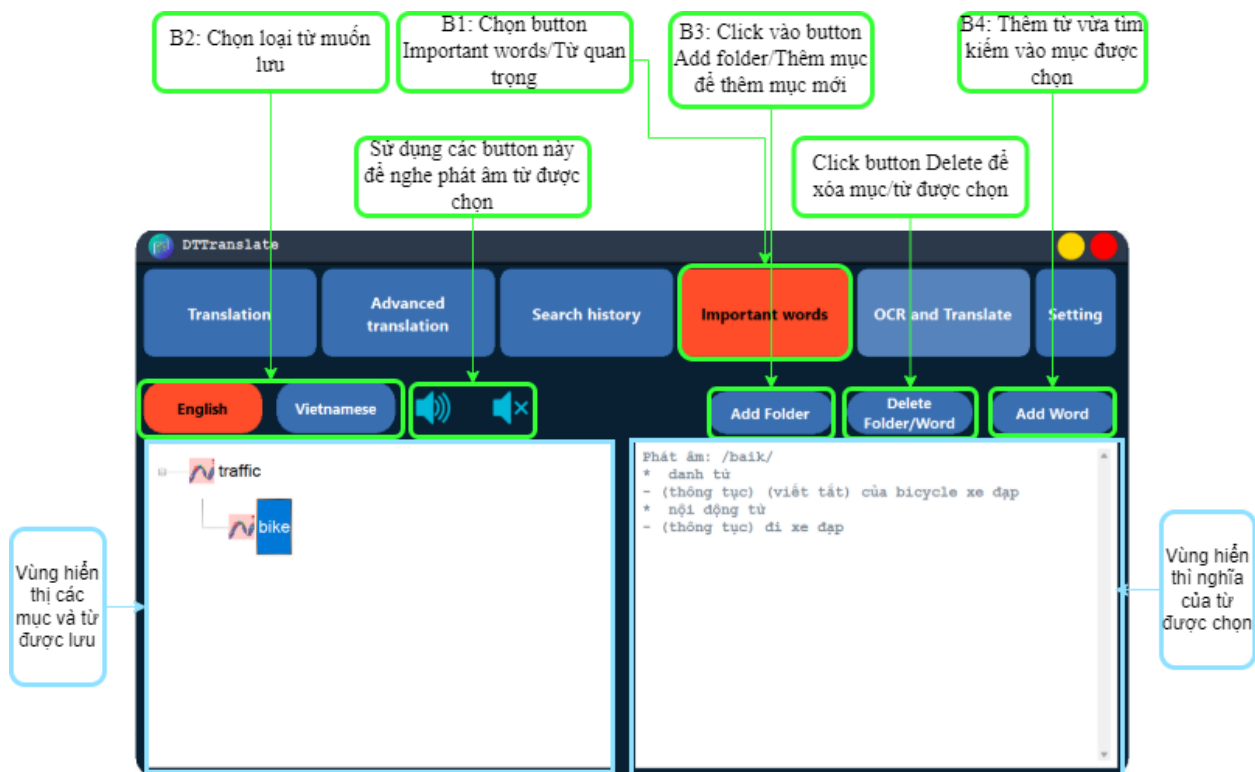
Hình 13: Cách sử dụng Search history/Lịch sử tìm kiếm

#### 4. Important words/Từ quan trọng

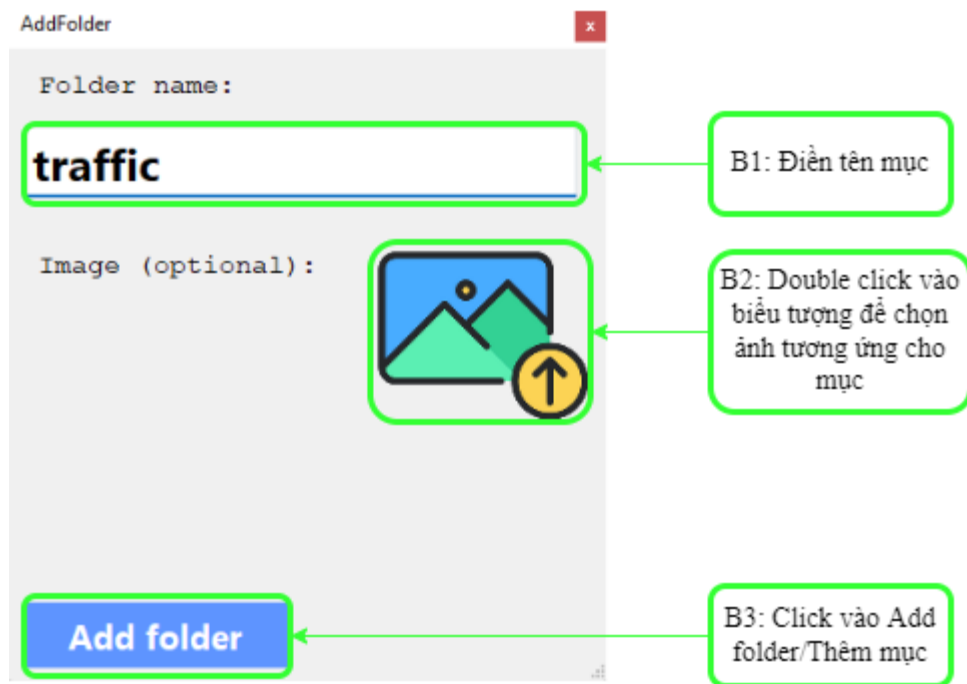
Cách sử dụng:

- Bước 1: Click vào Important words/Từ quan trọng để chuyển sang phần này.
- Bước 2: Chọn loại từ muốn lưu: Tiếng Anh, Tiếng Việt
- Bước 3: Click vào Add folder/Thêm mục để thêm mục lưu trữ mới. Sau khi click sẽ xuất hiện một form như Hình 15.
  - o Bước 3.1: Điền tên mục muốn tạo. Lưu ý tên mục muốn tạo không được trùng với mục đã tồn tại.
  - o Bước 3.2 (Tùy chọn): Click đúp vào biểu tượng hình ảnh để thêm hình ảnh đại diện cho mục
  - o Bước 3.3: Click vào Add folder/Thêm mục trong form để thêm mục
- Bước 4: Click vào mục muốn thêm sau đó click Add word/Thêm từ để thêm từ vừa tra ở phần Translation/Dịch vào mục được chọn. Kết quả xuất hiện như phần cây bên trái.

Nếu muốn xóa mục/từ, chọn mục/từ muốn xóa rồi nhấn Delete Folder/Word. Ngoài ra chọn từ muốn phát âm rồi click vào các biểu tượng hình Loa để nghe phát âm.



Hình 14: Cách sử dụng Important words/Từ quan trọng



Hình 15: Form sau khi click vào Add Folder/Thêm mục ở Hình 14

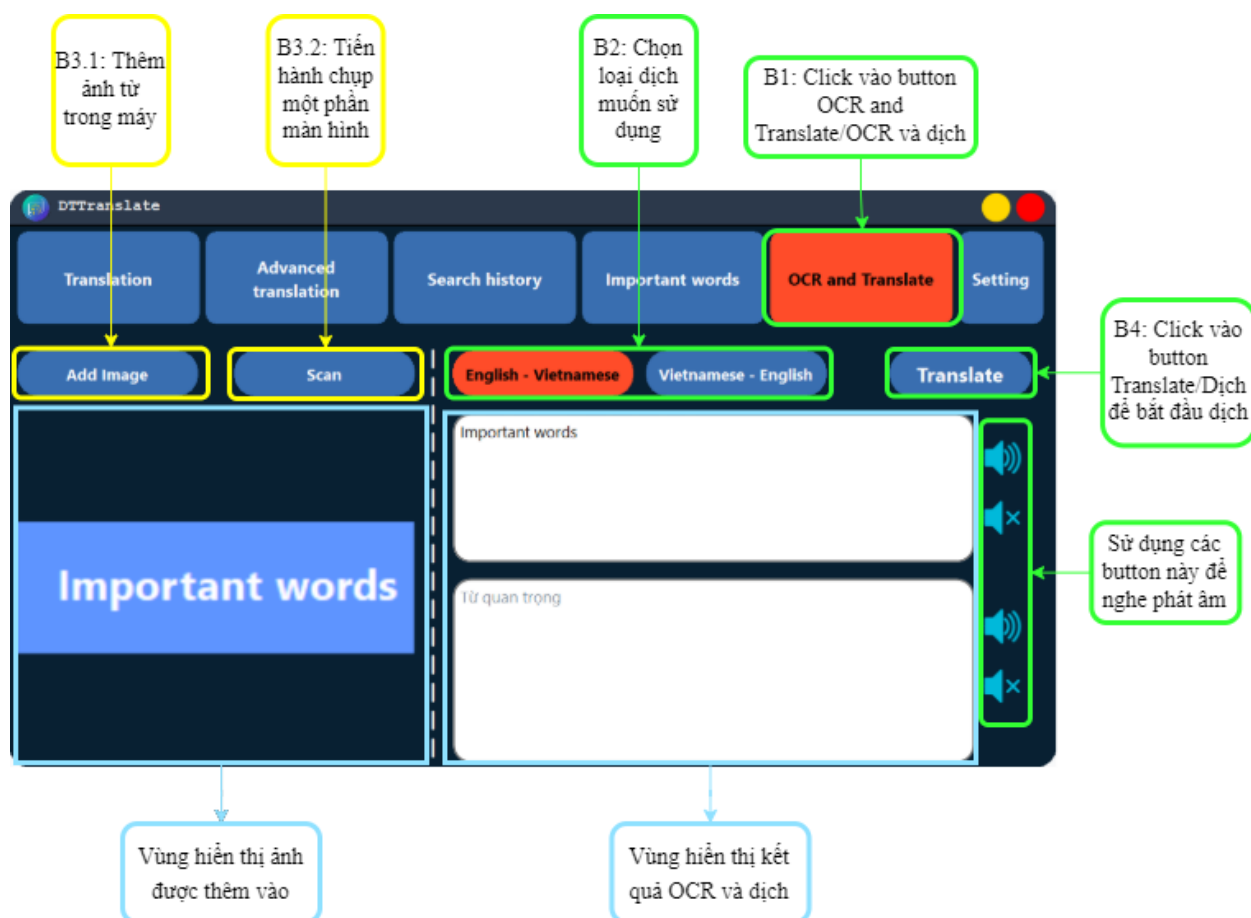
## 5. OCR and Translate/OCR và dịch

Phần OCR và Translate sẽ tìm kiếm chữ từ hình ảnh và dịch đoạn chữ đó.  
Cách dùng:

- Bước 1: Click vào OCR and Translate/OCR và dịch

- Bước 2: Chọn loại dịch muốn sử dụng: Anh – Việt, Việt – Anh
- Bước 3: Chọn nguồn hình ảnh. Người dùng có thể lấy hình ảnh từ máy tính bằng cách click vào Add image/Thêm ảnh hoặc chụp ảnh từ màn hình máy tính bằng Scan/Chụp màn hình.
- Bước 4: Sau khi thêm ảnh, click vào Translate/Dịch để tiến hành OCR và dịch.

Người dùng cũng có thể sử dụng các biểu tượng hình Loa bên phải để nghe phát âm của từng phần.



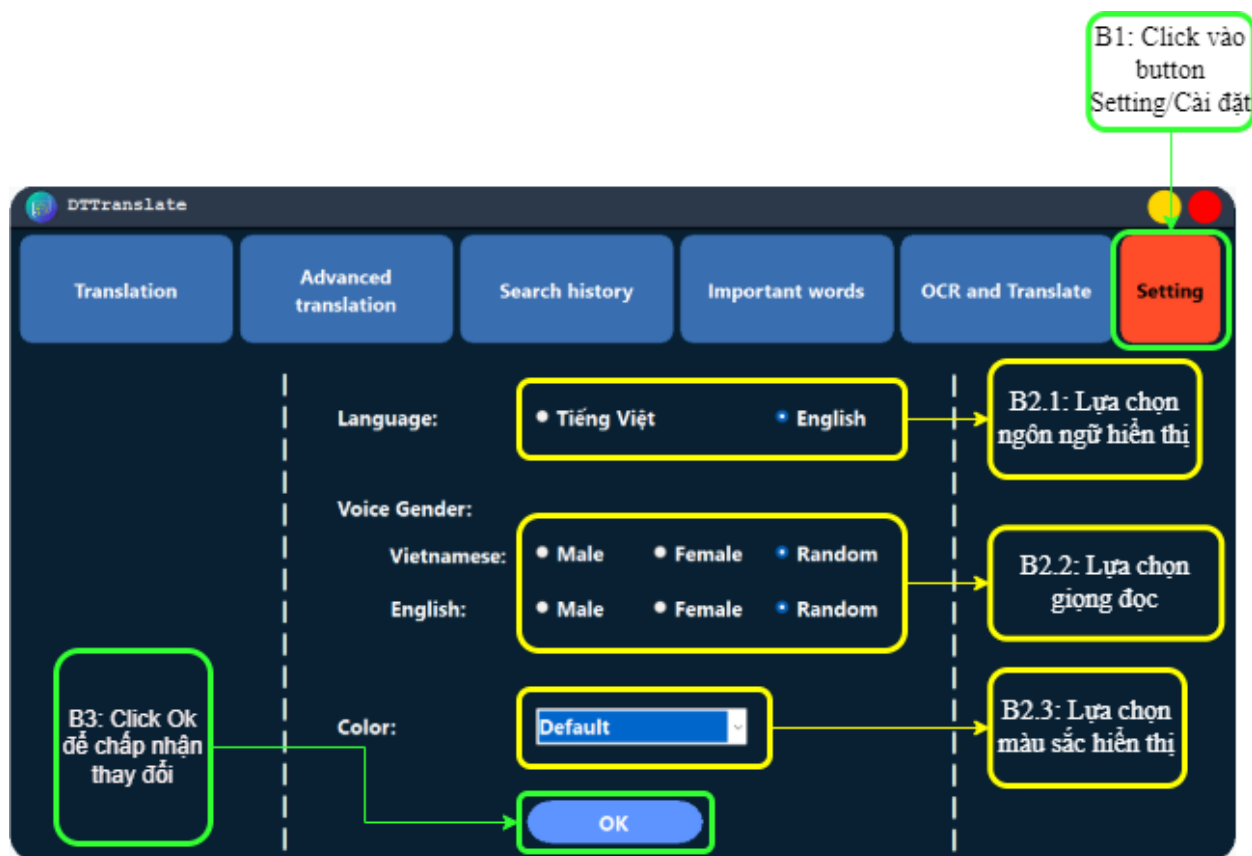
Hình 16: Cách sử dụng OCR and Translate/OCR và dịch

## 6. Setting

Người dùng click vào phần Setting/Cài đặt để chuyển sang phần Cài đặt. Ở đây người dùng có thể chọn Ngôn ngữ hiển thị, Giọng đọc cho từng tiếng, Màu sắc của giao diện.

Sau khi chọn xong, click vào OK để chấp nhận thay đổi.





Hình 17: Cách sử dụng Setting/Cài đặt

## V. Hướng phát triển

Vì phần mềm DTTranslate là một sản phẩm đồ án của môn Ngôn ngữ lập trình C# nên thời gian đầu tư chưa được nhiều như các sản phẩm khác đang có trên thị trường nên nhóm có một số định hướng có phát triển sản phẩm sau này như sau:

- Tối ưu lại code, thuật toán làm giảm thời gian thực hiện các chức năng có trong sản phẩm.
- Cài đặt thêm phần nhận diện giọng nói (speech to text).
- Mở rộng thêm phần trò chơi giúp học từ vựng.
- Bổ sung thêm dữ liệu từ điển.
- Và một số cải thiện phần giao diện nhằm thân thiện, đơn giản, hợp lý hơn.

## VI. Mô tả công việc của các thành viên trong nhóm và đề xuất

### 1. Công việc của các thành viên

| MSSV     | Sinh viên          | Công việc   | Nhận xét  |
|----------|--------------------|---|---|
| 19521388 | Hoàng Tiến Dũng    | <ul style="list-style-type: none"> <li>- Làm chức năng dịch cơ bản, dịch nâng cao, OCR và dịch, cài đặt, phát âm.</li> <li>- Tham gia báo cáo đồ án</li> <li>- Tham gia viết báo cáo đồ án</li> </ul> | Thực hiện tốt, hoàn thành trong thời gian đặt ra. |
| 19522410 | Nguyễn Thành Trọng | <ul style="list-style-type: none"> <li>- Làm phần giao diện, lịch sử tìm kiếm, lưu từ quan trọng.</li> <li>- Tham gia báo cáo đồ án</li> <li>- Tham gia viết báo cáo đồ án</li> </ul>                 | Thực hiện tốt, hoàn thành trong thời gian đặt ra. |

## 2. Đề xuất

Vì đây là môn học đầu tiên các thành viên nhóm em tiếp xúc với ngôn ngữ C#, nên có nhiều kiến thức chưa biết. Do đó, trong quá trình làm đã gặp nhiều kiến thức mới như lưu và sử dụng SQL từ C#, nhận diện chữ viết từ hình ảnh, sử dụng API và nhóm đã mất rất nhiều thời gian sửa lỗi và học các kiến thức mới.

Tuy sản phẩm chưa làm được phần mở rộng trò chơi học từ vựng như cô đã yêu cầu nhưng trong đó có những chức năng mới mà ít có trong các phần mềm PC có ở hiện tại như chụp ảnh từ màn hình, OCR ảnh, phát âm Việt-Anh. Nhóm em sẽ hoàn thiện sản phẩm tốt hơn trong tương lai.

Với độ hoàn thiện đồ án thì nhóm xin được đề xuất 9 điểm cho đồ án Phần mềm dịch thuật DTTranslate.

Nhóm chúng em xin chân thành cảm ơn!

## VII. Nguồn học và tham khảo

- [1] <https://www.howkteam.vn/>
- [2] <https://stackoverflow.com/>
- [3] <https://docs.microsoft.com/en-us/dotnet/api/>
- [4] <https://www.youtube.com/>