TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ Trial

**CHƯƠNG** **TRÌNH** **CHẤT** **LƯỢNG** **CAO**

**ĐỀ** **CƯƠNG** **KHÓA** **LUẬN** **TỐT** **NGHIỆP**

**Tên** **đề** **tài:**

**XÂY** **DỰNG** **CÔNG** **CỤ** **HỖ** **TRỢ** **DỊCH** **THUẬT**

**Họ** **tên,** **chức** **danh** **GVHD:** **PGS.TS.** **Đinh** **Điền**

**Sinh** **viên:** **Trần** **Minh** **Cảnh** (1653008) – **Nguyễn** **Thanh** **Long** (1653046)

**Thể** **loại** **KLTN/ĐATN:** **Nghiên** **cứu** **có** **ứng** **dụng** **demo**

**Thời** **gian** **thực** **hiện:** **Từ** 17/12/2019 **Đến** 31/7/2020

**Nội** **dung** **KLTN/ĐATN:**

Thế kỉ 21 là kỉ nguyên của bùng nổ thông tin và hội nhập toàn cầu. Cùng với sự phát triển của kinh tế, khoa học, văn hóa và sự bùng nổ trao đổi quốc tế, nhu cầu dịch thuật trong nhiều lĩnh vực khác nhau ngày càng lớn, đòi hỏi ngày càng nhiều dịch giả phải có khả năng dịch tài liệu, văn bản với số lượng khổng lồ trong nhiều lĩnh vực khác nhau. Với tình hình đó, công nghệ thông tin đã được áp dụng và tác động mạnh mẽ tới ngành dịch thuật với sự ra đời của Computer-assisted translation.

Computer-assistedtranslation, computer-aidedtranslationhayCATlà một phương pháp trong dịch thuật, dịch tài liệu trong đó người dịch sẽ sử dụng phần mềm máy tính để hỗ trợ và tạo điều kiện tốt nhất cho quá trình dịch thuật nhằm tăng năng xuất và chất lượng bản dịch. Đồng thời quản lý quá trình dịch thuật được kết quả tốt nhất.

Trong một khảo sát của Joanna Gough(2011) với hơn 224 dịch giả chuyên

nghiệp từ 42 quốc gia tham gia, đại đa số (94%) người dịch đã sử dụng các công cụ CAT trong đó 62% dịch giả đôi khi sử dụng và 32% dịch giả sử dụng thường

xuyên và chỉ có 6% tuyên bố rằng họ không sử dụng, thiểu số người không sử dụng với các lý do về : kinh phí, thời gian làm quen với CAT, nhu cầu sử dụng không nhiều. Qua khảo sát trên cho thấy CAT đã trở thành một phần tất yếu, hỗ trợ cho dịch giả rất nhiều.

Năm 2013, Proz đã có một cuộc khảo sát về phần mềm CAT được sử dụng nhiều nhất (có thể chọn nhiều phần mềm) với hơn 3000 dịch giả chuyên nghiệp tham gia. Qua đó thấy được các phần mềm sử dụng nhiều nhất là Trados(hơn 72%), memoQ (~26%), các phần mềm còn lại đều dưới 5%. Các phần mềm phổ biến trên đã rất tốt nhưng vẫn còn gặp nhiều bất cập như : giá cao, giao diện phức tạp, khó tiếp cận người dùng mới, một số tính năng cơ bản đôi khi có lỗi (ví dụ: lỗi xuất file của Trados, tốc độ truy cập TM khi làm online bị hạn chế của memoQ..). Nhận ra những điểm hạn chế đó cùng với mong muốn tìm hiểu rõ hơn về CAT đã tiếp thêm động lực cho chúng tôi nghiên cứu về đề tài này.

**Thông** **qua** **đề** **tài** **này** **:**

Chúng tôi được tìm hiểu công cụ hỗ trợ dịch thuật(CAT). Quá trình hình thành và chức năng của CAT, tìm hiểu cách thức hoạt động, kỹ thuật được sử dụng trong CAT. Chúng tôi đặt mục tiêu nghiên cứu về CAT một cách hiệu quả nhất, Tạo ra 1 ứng dụng CAT dễ sử dụng, thân thiện với người dùng mới, phù hợp cho các đối tượng không thuộc nhóm ngành công nghệ thông tin với các chức năng cơ bản:

- Hỗ trợ đa dạng file: txt, word, excel.

- So sánh độ tương đồng.

- Gióng hàng.

- Sử dụng Machine Translation.

Trial tôi sử dụng phương pháp tiếp cận: tìm kiếm, nghiên cứu tài liệu về CAT, sử dụng và khai thác các phần mềm CAT hiện nay, học hỏi những người đang nghiên cứu về CAT. Với cách thức nghiên cứu trên chúng tôi mong đợi sẽ đem lại kết quả tốt giúp chúng tôi hiều rõ và nắm vững kiến thức về CAT hơn,

tạo ra được một phần mềm CAT với các mục tiêu đã đề ra bên trên, góp phần phục vụ cho công tác dịch thuật của các dịch thuật viên.

**Các** **mốc** **thời** **gian** **nghiên** **cứu:**

|  |  |
| --- | --- |
| **THỜI** **GIAN** | **CÔNG** **VIỆC** |
| **17/12/2019** **-** **31/12/2019** | - **Tìm** **hiểu** **CAT** **là** **gì,** **quá** **trình** **phát** **triển,** **chức** **năng** **của** **CAT.** |
| **1/1/2020** **–** **15/1/2020** | - **Tìm** **hiểu** **về** **các** **phần** **mềm** **CAT,** **nghiên** **cứu** **về** **tính** **năng,** **điểm** **mạnh** **điểm** **yếu** **của** **từng** **phần** **mềm** **CAT** **hiện** **tại.**  - **Lưu** **vào** **sưu** **liệu** **để** **phục** **vụ** **việc** **viết** **khóa** **luận.** |
| **16/1/2020** **–** **30/1/2020** | - **Tìm** **hiểu** **cách** **thức** **hoạt** **động** **và** **công** **nghệ** **sử** **dụng** **trong** **CAT.**  - **Nghiên** **cứu** **mã** **nguồn** **mở** **của** **omegaT(phần** **mềm** **CAT** **miễn** **phí).** |
| **1/2/2020** **–** **15/2/2020** | - **Tìm** **hiểu** **về** **các** **ngôn** **ngữ** **lập** **trình** **CAT,** **các** **phần** **mềm** **lập** **trình** **nên** **sử** **dụng** **.** |
| **16/2/2020** **–** **30/2/2020** | - **Thực** **hiện** **làm** **demo** **phần** **mềm** **CAT** **(input,** **output** **file:** **txt).** |
| **1/3/2020** **–** **15/3/2020** | - **Nghiên** **cứu** **và** **tìm** **hiểu** **cách** **đọc** **và** **xuất** **các** **loại** **file** **văn** **bản** **khác** **trên** **CAT** **(vẫn** **giữ** **nguyên** **đặc** **điểm** **của** **file** **gốc).**  - **Cải** **thiện** **giao** **diện** **demo.** |
| **16/3/2020** **–** **31/3/2020** | - **Nhận** **phản** **hồi** **từ** **giáo** **viên** **hướng** **dẫn** **và** **cải** **thiện** **demo.** |
| **1/4/2020** **–** **30/4/2020** | - **Phát** **triển** **demo** **phần** **mềm** **CAT** **với** **các** **chức** **năng:** **sử** **dụng** **gióng** **hàng,** **so** **sánh** **tương** **đồng** **các** **cấp** **độ,** **sử** **dụng** **machine** **translation.** |
| **1/5/2020** **–** **16/5/2020** | - **Tiếp** **tục** **đọc** **và** **tìm** **hiểu** **các** **tài** **liệu** **để** **có** **thêm** **sưu** **liệu** **cho** **luận** **văn.**  - **Nhận** **phản** **hồi** **demo** **và** **cải** **thiện.** |

|  |  |
| --- | --- |
| **30/5/2020** **–** **15/6/2020** | - **Cập** **nhật** **thêm** **vào** **luận** **văn** **với** **các** **sưu** **liệu** **và** **kiến** **thức** **đã** **có** **được** **trong** **thời** **gian** **làm** **việc.** |
| **16/6/2020** **–** **30/6/2020** | - **Kiểm** **tra,** **sửa** **lỗi** **và** **hoàn** **thiện** **chương** **trình** **CAT.** |
| **1/7/2020** **–** **31/7/2020** | - **Kiểm** **tra,** **sửa** **lỗi** **và** **hoàn** **hiện** **luận** **văn.** |

**Ý** **kiến** **của** **giảng** **viên** **hướng** **dẫn**

***Chữ*** ***ký*** ***của*** ***giảng*** ***viên*** ***hướng*** ***dẫn***

**TP.** **HCM,** **08/03/2020**

***Chữ*** ***ký*** ***(các)*** ***sinh*** ***viên***