Hướng dẫn sử dụng CAT

Tủ sách Nhất Nghệ Tinh

Phiên bản: 4.0
Ngày: 11.4.2019
Thực hiện: Doãn Minh Đăng
Trần Dương Hân
Hiệu đính: Phạm Ngọc Thúy

Mục lục

1 Lời nói đầu 2
2 Giới thiệu CAT 3
3 CAT dùng trong Tủ sách NNT 4
3.1 Các phần mềm sử dụng 4
3.2 Tải về và cài đặt các phần mềm sử dụng 4
3.3 Tạo tài khoản ở Gitlab.com 4
3.4 Ban Quản trị CAT của Tủ sách NNT 5
4 Các hướng dẫn chung 6
5 Mô hình dịch nhóm 10
6 Tiến trình dịch nhóm 11
7 Hướng dẫn sử dụng CAT cho nhóm viên (dịch hay hiệu đính) 12
7.1 Nhiệm vụ của nhóm viên 12
7.2 Các dữ liệu đầu vào từ trưởng nhóm 13
7.3 Các bước thực hiện 14
7.4 Các câu hỏi thường gặp 16

1 Lời nói đầu

Tài liệu này được viết để hỗ trợ các thành viên trong Tủ sách Nhất Nghệ Tinh (Tủ sách NNT) tham gia việc dịch thuật theo nhóm thông qua công cụ CAT. Quy trình dịch nhóm với CAT được phổ biến ở khóa học về CAT vào đầu tháng 4.2018 ở Offenbach, ngay sau đó được thử nghiệm với dự án dịch sách Chemietechnik, và ôn lại vào khóa học CAT tháng 4.2019 ở Bingen.

Trong quá trình dịch thuật các cuốn sách học và dạy nghề trước mắt cũng như trong tương lai, tài liệu này sẽ được cập nhật thường xuyên và gởi đến các bạn qua Tủ sách để kết quả việc dịch thuật của chúng ta ngày càng nhanh và chính xác hơn.

Mọi thắc mắc, góp ý xin gởi về:

Doãn Minh Đăng doanminhdang@gmail.com
Trần Dương Hân tranhan11@gmail.com

2 Giới thiệu CAT

CAT là viết tắt của cụm từ *Computer-Assisted Translation* hay *Computer-Aidded Translation* và mang ý nghĩa "hỗ trợ dịch thuật thông qua phần mềm".

Một công cụ (*tool*) CAT bao gồm:

- Bảng thuật ngữ (*Glossary*)
- Bộ nhớ dịch (*Translation Memory*)
- Tự điển (*Dictionary*)
- Tích hợp với một hay nhiều phần mềm dịch máy (*Machine Translation*, MT)
- Các phần mềm chuyển đổi định dạng văn bản (*Conversion programs*)
- Các công cụ để tạo và ghi vào bộ nhớ dịch các văn bản gốc và văn bản dịch đã được thực hiện từ trước
- Phần mềm theo dõi và hỗ trợ tiến trình dịch thuật.

Trong khi việc dịch máy (MT) chỉ là việc chuyển đổi một từ, một câu hay một văn bản từ một ngôn ngữ này sang một ngôn ngữ khác mà không có sự tham gia trực tiếp của con người thì CAT hỗ trợ việc dịch thuật dựa trên những thành quả đã đạt được qua các văn bản đã được dịch trước. Trong quá trình hỗ trợ, người dùng tiếp tục cập nhật các bảng thuật ngữ, bộ nhớ dịch hay tự điển để việc dịch thuật càng ngày càng nhanh và chính xác hơn. Mặt khác, CAT cũng hỗ trợ việc tích hợp (*integration*) các phần mềm dịch máy để tạo ra các tài liệu thô ban đầu (*raw document*) nhằm mục đích gia tăng tốc độ dịch thuật.

Ngoài ra, việc dùng bảng thuật ngữ đã tạo điều kiện tiên quyết để đạt được việc dùng thuật ngữ thống nhất trong một hay nhiều văn bản nếu mang cùng một ý nghĩa. Do đó, cùng với phần mềm theo dõi và hỗ trợ tiến trình dịch, CAT hoàn toàn thích hợp cho việc dịch theo nhóm hay nói gọn là dịch nhóm (*Team Translation*).

3 CAT dùng trong Tủ sách NNT

Nhận thấy sự cần thiết phải gia tăng tốc độ dịch thuật, dễ dàng đạt được sự thống nhất trong việc sử dụng thuật ngữ trong một hay nhiều văn bản, Tủ sách Nhất Nghệ Tinh (Tủ sách NNT) đã quyết định dùng CAT lần đầu tiên cho dự án dịch cuốn sách "Kỹ thuật Hóa học" (*Chemietechnik*) và đã tổ chức một khóa học về CAT ở Offenbach vào cuối tuần 6-8.4.2018 dưới sự hướng dẫn của bạn trẻ Doãn Minh Đăng và với sự tham dự của 21 thành viên khác của Tủ sách.

3.1 Các phần mềm sử dụng

- **JRE (Java Runtime Environment)**: Đây là phần mềm cơ sở, là nền tảng để chạy các phần mềm khác dùng trong CAT vì các phần mềm này được viết với ngôn ngữ Java.

- **OmegaT**: Đây là phần mềm chính của CAT dùng trong Tủ sách NNT nhằm mục đích hỗ trợ việc dịch nhóm. Ở đây các tài nguyên của dự án dịch như bảng thuật ngữ, bộ nhớ dịch được chia sẻ qua mạng và các nhóm viên có thể làm việc cùng lúc trên nhiều máy tính khác nhau. Để đạt được chức năng này, OmegaT cần một kho lưu trữ trên mạng để quản lý các tài nguyên của dự án dịch.

- **Gitlab.com**: Tủ sách NNT dùng trang mạng này làm kho lưu trữ (*repository*) cho các dự án dịch. Trang mạng này dùng GIT để quản lý dự án dịch nghĩa là sẽ tự động tạo ra các bản sao của dự án dịch trên từng máy tính của các nhóm viên và họ làm việc với OmegaT như đang dịch một văn bản ở máy tính cá nhân.

3.2 Tải về và cài đặt các phần mềm sử dụng

JRE (Java Runtime Environment)

- Tải về: https://java.com/en/download
- Cài đặt: Khởi động (*launch*) tập tin jre-8u161-windows-x64.exe vừa tải về để cài đặt.

OmegaT

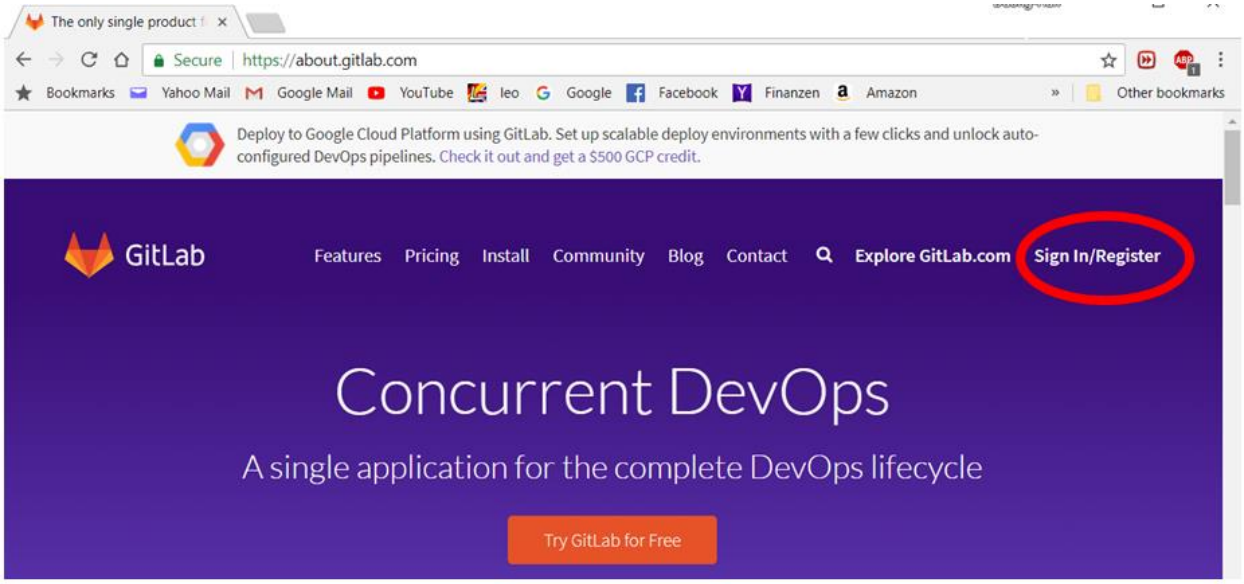
- Tải về: http://omegat.org/download và chọn phiên bản "Latest Version" (cần dùng phiên bản **4.1.5 Update 4 hoặc mới hơn**, không dùng phiên bản có số hiệu 3.x).
- Cài đặt: Không cần cài đặt mà chỉ cần giải nén (unzip) vào một thư mục (folder) tự chọn.

3.3 Tạo tài khoản ở Gitlab.com

Mỗi thành viên tham gia dịch nhóm dùng CAT cần tạo một tài khoản ở Gitlab.com vì chúng ta đã quyết định dùng trang mạng này làm kho lưu trữ (repository). Theo hướng dẫn các bước sau để tạo một tài khoản của bạn ở Gitliab.

1. Dùng trình duyệt Firefox, Chrome, … để vào trang https://about.gitlab.com/
2. Nhấn *Sign In/Register* (Đăng nhập/Đăng ký)
3. Trên trang mới, nhấn *Register* (Đăng ký)

4. Cho vào các thông tin đòi hỏi:
    - *Full name* (Họ và tên)
    - *Username* (Tên người dùng)
    - *Email* và *Email confirmation* (2x Địa chỉ email để tránh gõ nhầm)
    - *Password* (Mật khẩu) ít nhất 8 ký tự
    - Đánh dấu vào ô "*Ich bin kein Roboter*" (cần thiết để chống việc phá trang mạng)
    - Nhấn vào *Register*
5. Gitlab sẽ gửi email để xác nhận việc tạo tài khoản, cần làm theo hướng dẫn trong đó để xác nhận.
6. Sau khi nhận được thư của Gitlab về việc đăng ký thành công, trưởng nhóm cũng như các nhóm viên phải liên lạc với BQL CAT của Tủ sách NNT để được ghi tên vào nhóm (*group*) UBTT-NNT. Bạn cần thông báo cho trưởng nhóm của bạn hay Ban Quản trị (BQT) Gitlab của Tủ sách NNT "Tên người dùng" của bạn. Bạn sẽ được ghi tên vào nhóm dịch UBTT-NNT ở Gitlab và tùy theo công việc mà bạn sẽ được xếp vào thành viên với một trong những vai trò: khách, nhà phát triển hay chủ (guest, developer, master).



3.4 Ban Quản trị CAT của Tủ sách NNT

Khi có vấn đề hay thắc mắc về việc sử dụng CAT các bạn hãy liên lạc trực tiếp với trưởng nhóm hoặc gởi thư đến cả ba người trong Ban Quản trị (BQT) CAT của Tủ sách NNT. Việc gởi đến cho cả ba người là cần thiết để tránh trường hợp người bạn gởi đến đang ở nơi không có mạng và hơn nữa cũng để BQT luôn biết các vấn đề nảy sinh trong lúc sử dụng và cũng để luôn cập nhật chính tài liệu này.

Doãn Minh Đăng doanminhdang@gmail.com
Trương Ngọc Giao tngiao@gmail.com
Trần Dương Hân tranhan11@gmail.com

4 Các hướng dẫn chung

Giải nén khi dùng Windows 10

Khi dùng Windows 10 và nếu bạn nhấn đúp vào tập tin đã được nén (đuôi của tên tập tin là zip), Windows 10 sẽ hiển thị một cửa sổ với nội dung là các tập tin bên trong của tập tin nén. Tuy vậy bạn rất dễ lầm với cửa sổ chứa thư mục của các tập tin đã được giải nén và khi nhấn đúp một tập tin nào đó trong "thư mục" này thì một hộp thoại (dialog box) sẽ hiện ra với cậu hỏi bạn muốn giải nén không.

Để tránh trường hợp lầm lẫn này (đã giải nén rồi mà Windows cứ hỏi hoài là muốn giải nén không!), bạn nên xóa tập tin nén hay dời đi một nơi nào khác sau khi giải nén!

Gitlab

Trang mạng Gitlab là kho lưu trữ (*repository*) cho các dự án dịch sách của Tủ sách NNT. Nhóm UBTT-NNT với địa chỉ https://gitlab.com/UBTT-NNT đã được tạo ra và hiện đã có vài dự án thử nghiệm trong nhóm này. Địa chỉ này được ẩn để bảo mật.

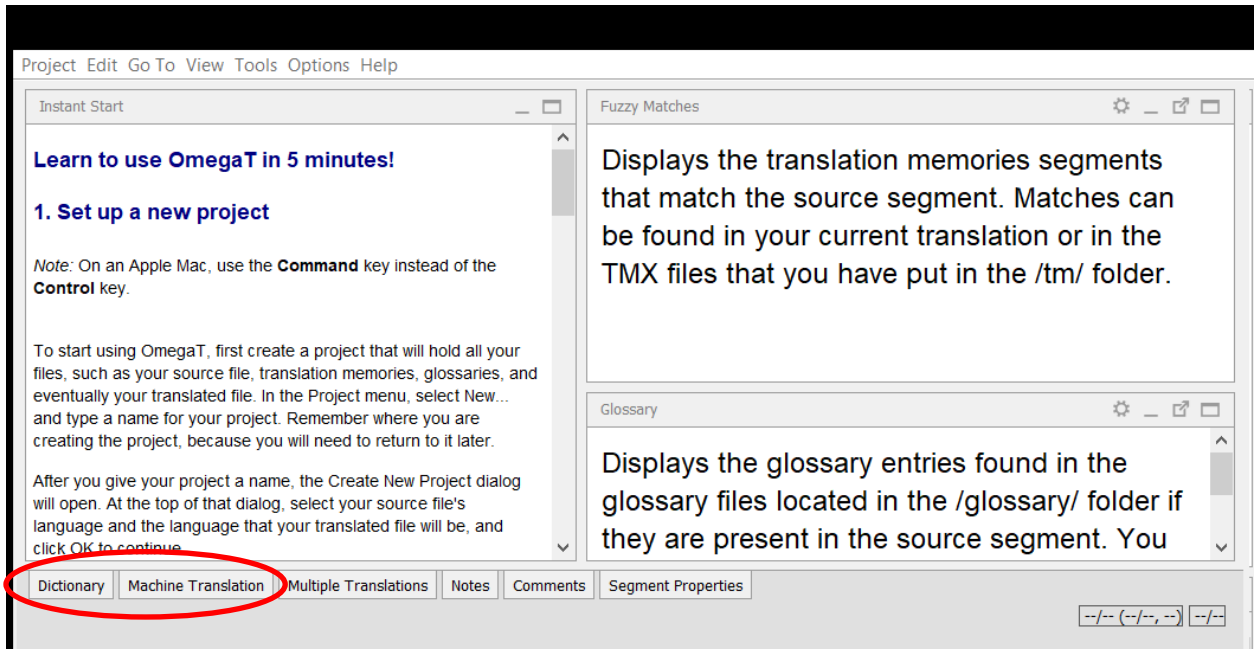
OmegaT

Khởi động OmegaT: Vào thư mục đã giải nén của phần mềm OmegaT và nhấn đúp vào tập tin *OmegaT.bat*. Nếu máy tính bạn không hiển thị phần đuôi của tập tin, bạn hãy chọn tập tin *OmegaT* với *Type* là *Windows Batch File*. Nếu dùng Windows 10, bạn có thể chỉnh lại để phần đuôi tập tin được hiển thị bằng cách mở một thư mục bất kỳ, chọn *View ở* thanh menu *(menu bar*) và sau đó làm dấu ô "*File name extensions*".

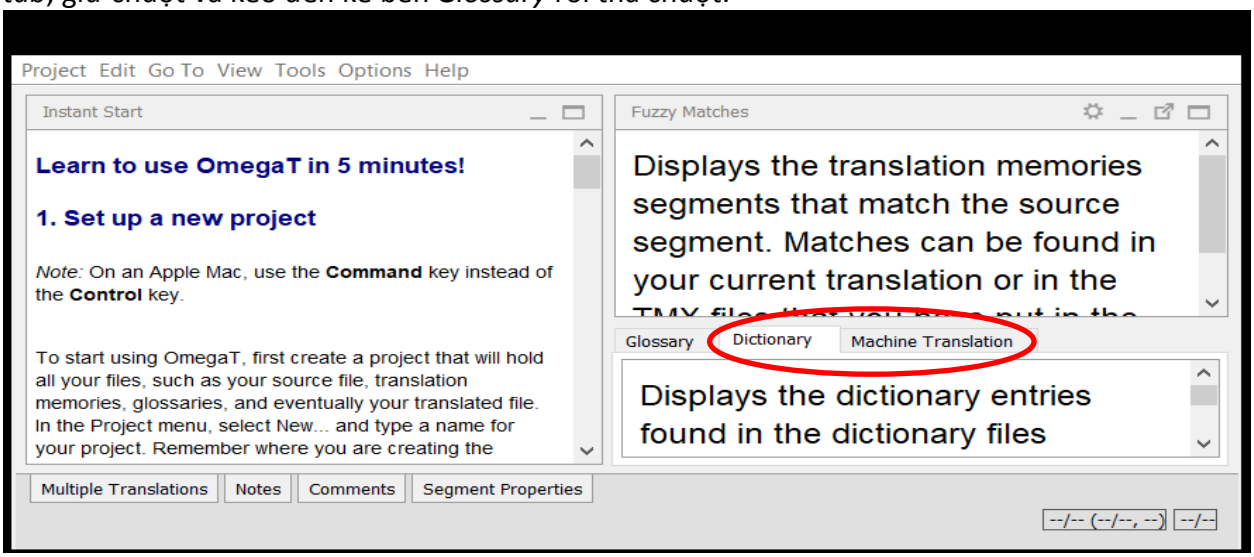
Những tùy chọn hữu ích khi dùng OmegaT

Dời các tab thường xuyên sử dụng

Phía dưới cùng của giao diện (User Interface) OmegaT có một số *tab* như *Dictionary, Machine Translation*, … mà nội dung chỉ hiện ra khi ta bấm vào đó. Vì *Dictionary* và *Machine Translation* được dùng tương đối thường xuyên nên sẽ rất hữu ích nếu ta đưa hai *tab* này vào chung với *Glossary ở* phần dưới bên phải của giao diện.

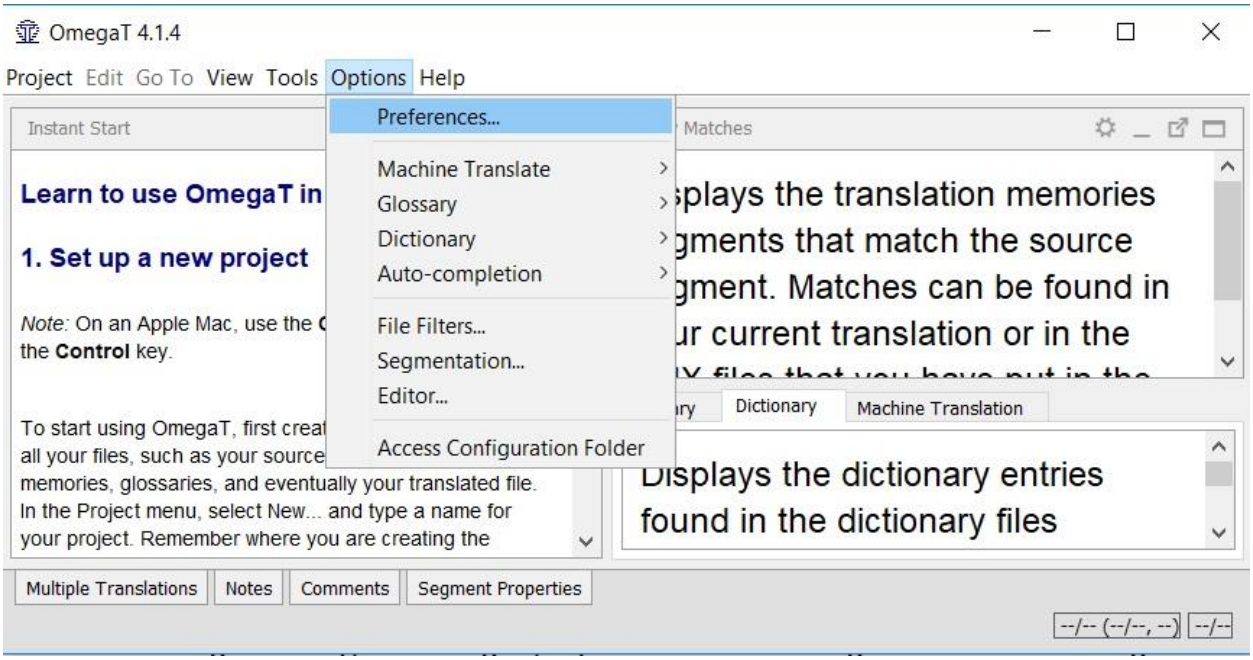


Dời *tab* bằng cách bấm chuột vào *tab* để hiển thị nội dung rồi bấm vào thanh ngang có tên của *tab*, giữ chuột và kéo đến kế bên *Glossary* rồi thả chuột.

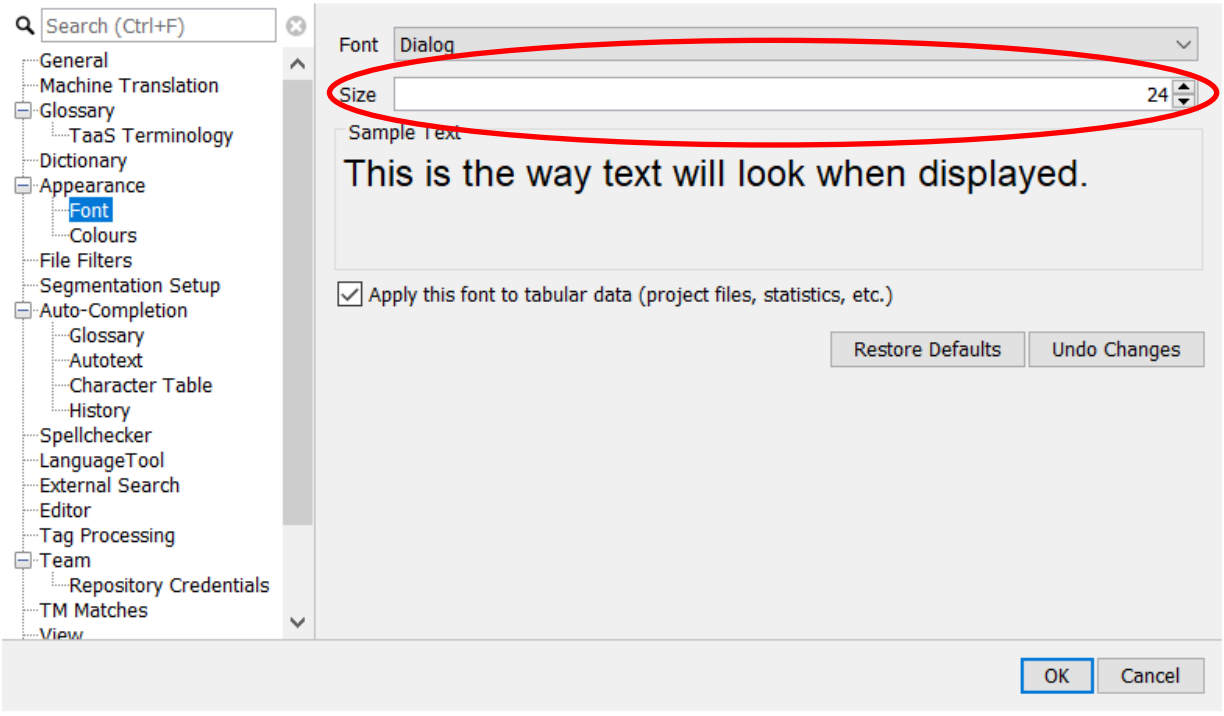


Làm lớn cỡ chữ trong giao diện

Bấm *Menu -> Options -> Preferences*

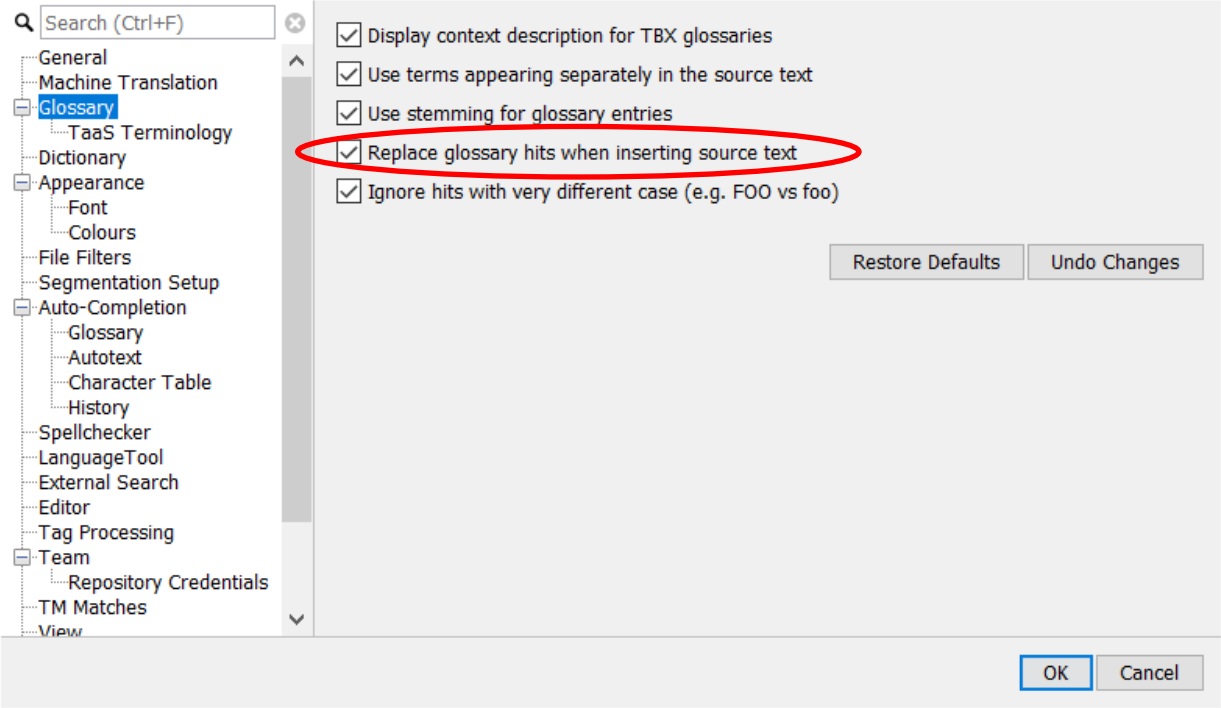


Sau đó bấm *Font* và chọn *Size 24*



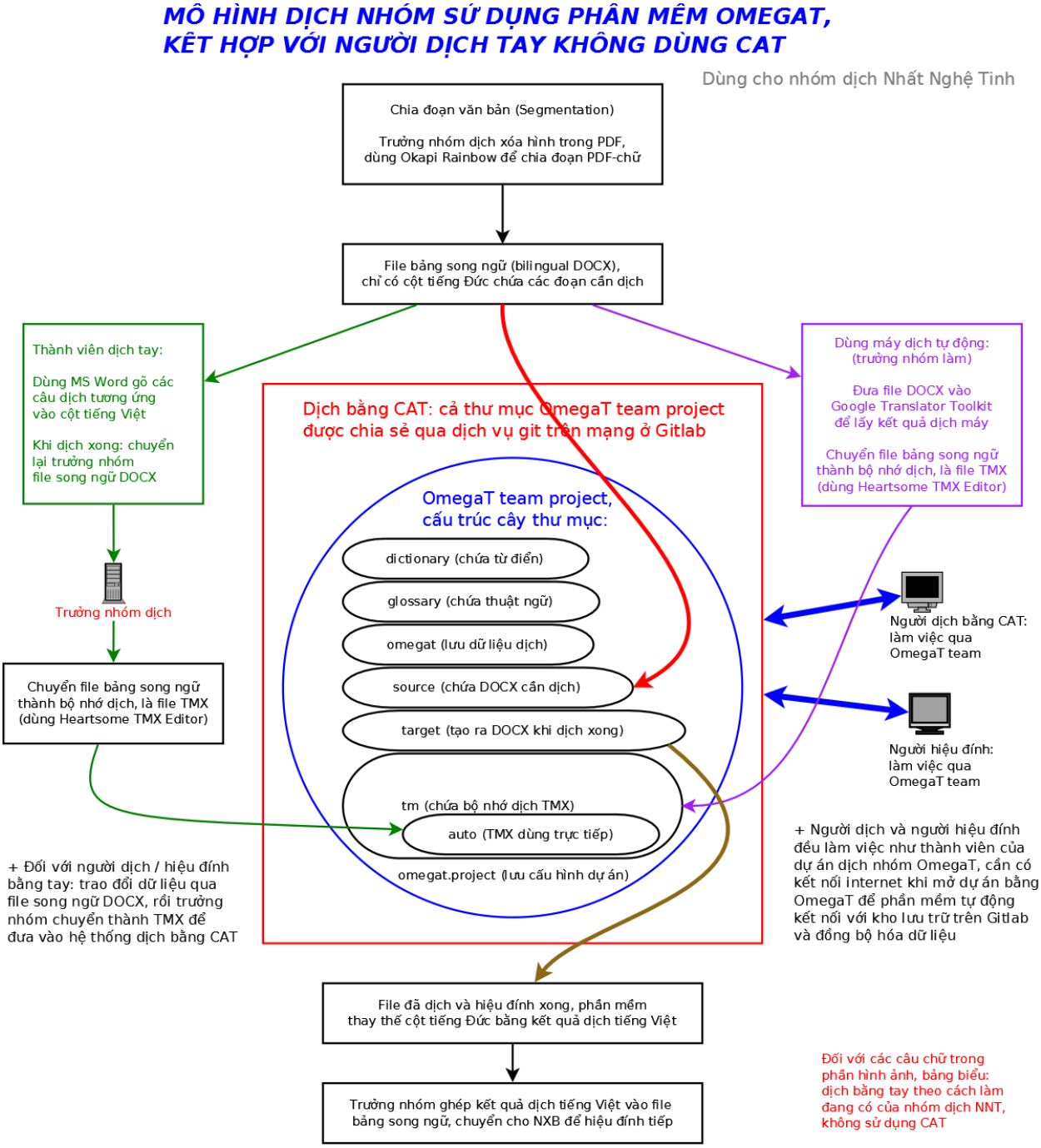
Tự động dịch các từ có trong Glossary (bảng thuật ngữ)

Bấm *Menu -> Options -> Preferences* như trên. Sau đó chọn *Glossary* và làm dấu (*mark*) ở *Replace glossary hits when inserting source text*.



Lựa chọn *Replace glossary hits when inserting source text* có tác dụng như sau: khi OmegaT mở một đoạn để dịch, phần mềm sẽ kiểm tra xem trong câu tiếng Đức đó có những từ (thuật ngữ) nào đã có lưu ở dữ liệu Glossary, và tự động thay từ đó bằng từ tiếng Việt trong Glossary để tạo thành câu văn dịch ban đầu, cho người dịch tiện sửa tiếp. Nếu không có lựa chọn này, thì câu văn dịch ban đầu là sao chép y hệt câu văn tiếng Đức, và người dịch cần xem thêm Glossary để biết từ tiếng Việt tương ứng và gõ vào câu văn dịch.

5 Mô hình dịch nhóm



Trong mô hình này, trưởng nhóm thực hiện hầu hết các công đoạn chuẩn bị: chia nhỏ tập tin PDF, xóa hình và bảng trong các tập tin PDF để phân đoạn văn bản, tập hợp thuật ngữ (thu thập trong Tủ sách NNT) và bộ nhớ dịch (có thể dùng Google Translator Toolkit để tạo bộ nhớ dịch tham khảo từ máy dịch tự động), rồi tạo dự án OmegaT và đưa lên Gitlab. Nhóm viên chỉ dùng OmegaT (có nối mạng internet) để dịch.

6 Tiến trình dịch nhóm

Phần mềm: PDF Split and Merge Basic (2), Okapi (3), Word (4), OmegaT (5, 7, 8, 9), Git tools cho trưởng nhóm (6). **Nhóm viên: chỉ làm bước 7, 8.**

Công cụ hỗ trợ: tạo Glossary; LF Aligner; xử lý TMX với Heartsome TMX Editor

7 Hướng dẫn sử dụng CAT cho nhóm viên (dịch hay hiệu đính)

7.1 Nhiệm vụ của nhóm viên

Nhiệm vụ của nhóm viên khi tham gia dịch bằng CAT là:

1. Tạo một tài khoản ở trang web https://gitlab.com, thông báo username cho BQT CAT của UBTT để được đưa vào nhóm https://gitlab.com/UBTT-NNT
2. Cài đặt phần mềm JRE (Java Runtime Environment) và OmegaT trên máy tính
3. Dùng OmegaT để tải dự án dịch nhóm về máy tính, và tiến hành dịch trên OmegaT.

7.2 Các dữ liệu đầu vào từ trưởng nhóm

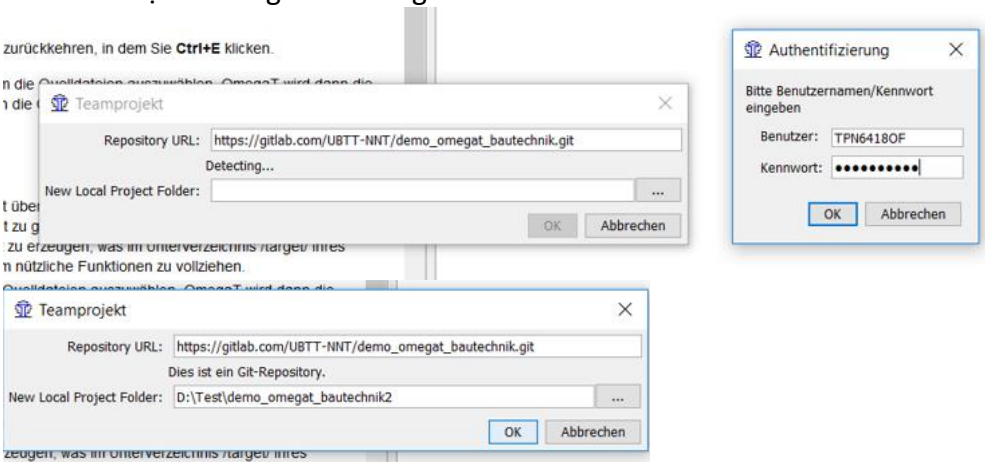
Trưởng nhóm sẽ gửi cho nhóm viên địa chỉ kho lưu trữ dự án dịch nhóm (repository) để tham gia dịch/hiệu đính, có dạng: https://gitlab.com/UBTT-NNT/Chemietechnik_1.git (kết thúc là .git). Trưởng nhóm cũng cho biết mỗi thành viên cần dịch/hiệu đính trên file tên gì.

Kho lưu trữ đó chứa toàn bộ các dữ liệu cần thiết để dịch với CAT, ngoài ra các tập tin PDF sẽ được trưởng nhóm gửi thêm để nhóm viên đánh số vào tập tin PDF khi dịch chữ trong hình và bảng, và trình bày bản dịch phần hình ảnh bằng tay.

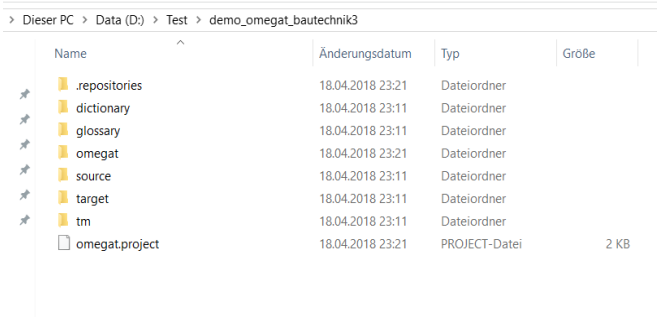
7.3 Các bước thực hiện

1. Mở phần mềm OmegaT, chọn menu **Projekt** -> **Teamprojekt herunterladen…**
   (Lưu ý: máy tính cần nối mạng internet)
   Điền vào địa chỉ repository mà mình đã nhận từ trưởng nhóm, gõ Enter.

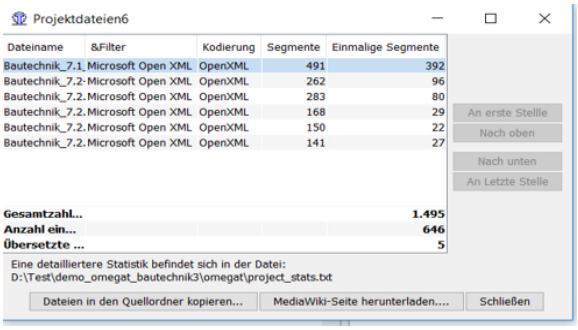
2. Phần mềm OmegaT sẽ hỏi thông tin đăng nhập, người dùng cần điền vào tên đăng nhập (username) và mật khẩu (password) của mình ở server gitlab.com. Phần mềm chỉ hỏi 1 lần và lưu lại để dùng cho những lần sau.

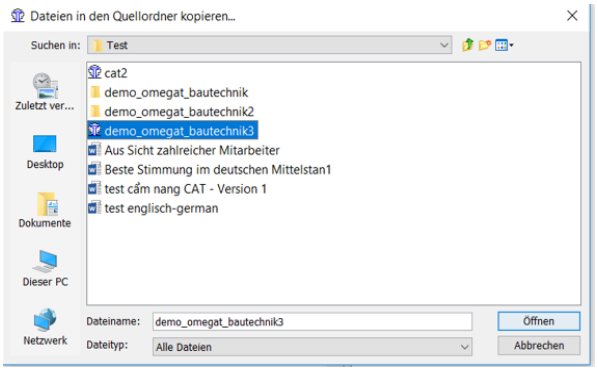


3. Sau khi điền các thông tin thành công, OmegaT sẽ hỏi mình muốn lưu dự án này vào chỗ nào trong máy tính cá nhân (PC). Chọn đường dẫn mình muốn lưu dự án, OmegaT sẽ tải về máy tính của mình dự án dịch nhóm đó. **Lưu ý**: thư mục nơi lưu dự án phải là chưa tồn tại (OmegaT sẽ tạo mới thư mục này). Sau này mỗi khi làm việc, chỉ cần mở lại dự án đã lưu trong máy tính.

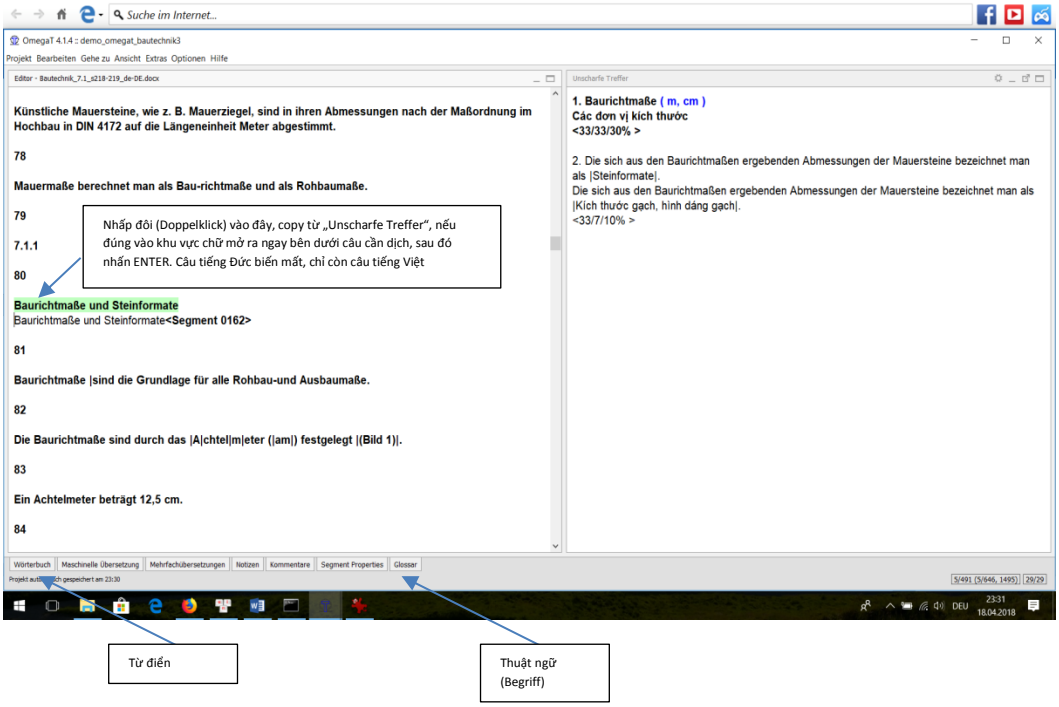


4. Sau khi OmegaT đã tải dự án về PC, hoặc mỗi khi dùng OmegaT để mở lại dự án đó: chọn tập tin (Datei) mà mình được giao làm việc, để OmegaT chuyển đến tập tin đó, rồi đóng cửa sổ chọn tập tin lại.





5. Trong cửa sổ làm việc của OmegaT: nhấp đúp vào từng đoạn cần dịch, gõ nội dung dịch vào khu vực chữ mở ra ngay bên dưới đoạn cần dịch. Mỗi khi chọn sang đoạn mới, thì phần dịch của đoạn cũ sẽ được tự động lưu lại. Phần mềm OmegaT thường xuyên kết nối internet để tự động "sync" (đồng bộ hóa) dữ liệu dịch trong PC với kho lưu trữ (repository) ở gitlab.com.

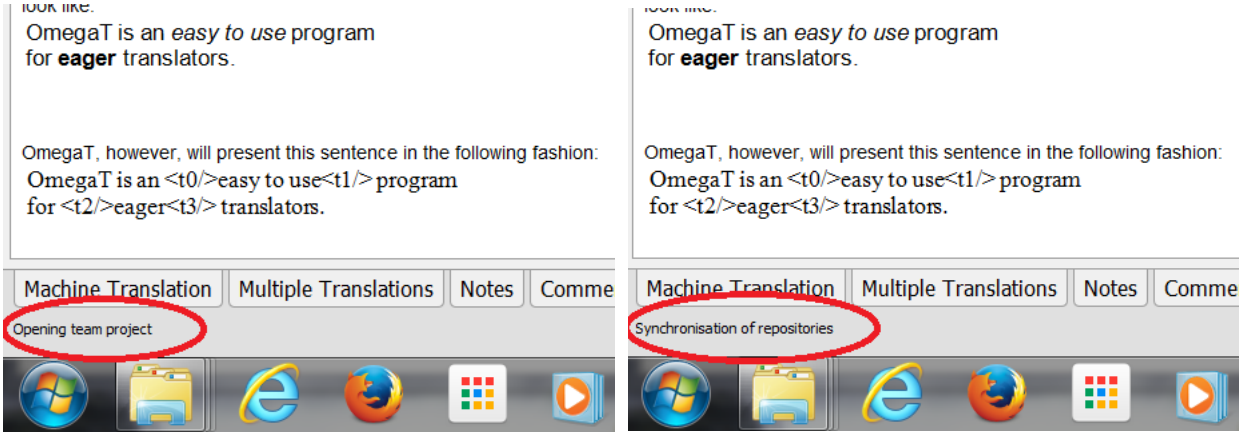


7.4 Các câu hỏi thường gặp

CÂU HỎI: Tôi không tải được dữ liệu từ kho lưu trữ khi khai báo địa chỉ dự án dịch nhóm.

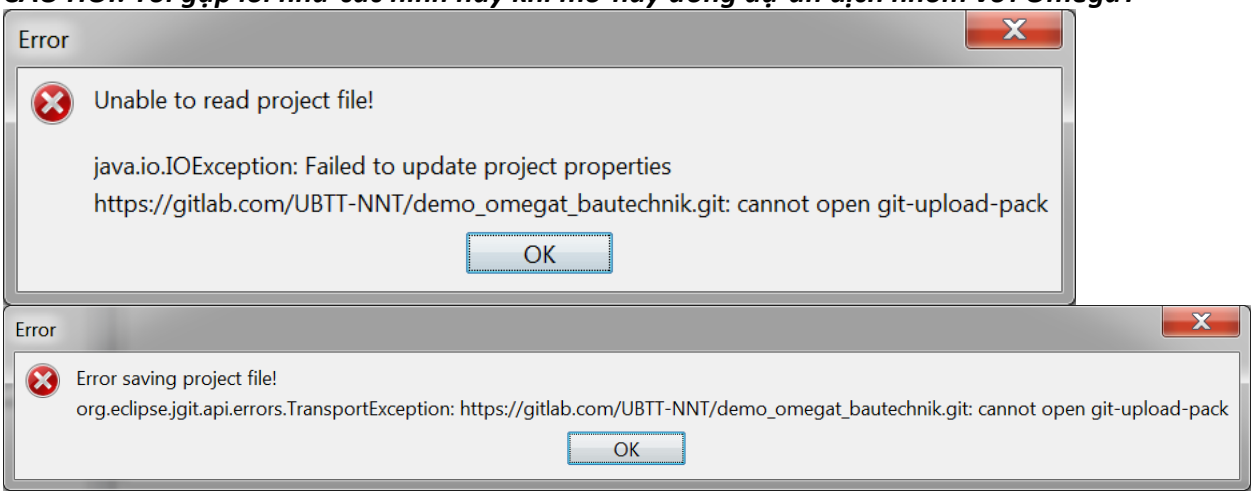
TRẢ LỜI: Có thể do tài khoản của nhóm viên chưa được đưa vào nhóm UBTT-NNT và đặt quyền hạn ở dự án. Cần trao đổi lại với trưởng nhóm.

CÂU HỎI: Dự án trong PC không được "sync" (đồng bộ hóa) tự động với kho lưu trữ (cách nhận biết: Khi dùng OmegaT mở dự án dịch nhóm, bình thường OmegaT sẽ dừng một thời gian và hiện ra thông báo ngắn ở bên dưới cửa sổ OmegaT là "Opening team project" và đang "Synchronisation of repositories", nếu không có giai đoạn chờ này là dự án không được tự động "sync" với kho lưu trữ)



TRẢ LỜI: Nếu không thấy quá trình "sync" hiện ra khi mở dự án dịch nhóm bằng OmegaT, thì có khả năng mối liên kết giữa PC và Gitlab đã bị OmegaT xóa do một lỗi nào đó. Cách giải quyết: hỏi trưởng nhóm, trưởng nhóm sẽ gửi cho tập tin *omegat.project* có lưu trữ thông tin về mối liên kết giữa PC và Gitlab, rồi lấy tập tin này thay cho tập tin *omegat.project* trong thư mục chứa dự án OmegaT trên PC.

CÂU HỎI: Tôi gặp lỗi như các hình này khi mở hay đóng dự án dịch nhóm với OmegaT



TRẢ LỜI: Đó là do máy tính không có kết nối internet, nên OmegaT không thể liên lạc với kho lưu trữ để đồng bộ hóa dữ liệu với công cụ git. Khi dùng OmegaT để dịch nhóm, máy tính cần nối mạng để dữ liệu dịch trong máy tính được “sync” với dữ liệu trong kho lưu trữ ở Gitlab vào thời điểm mở hay đóng dự án dịch nhóm.