



Terminology Management (Termbase)

V.20220428



Key Technology Tools

Utilising Glossary Converter and EXTRACT on remote PC

Glossary Converter

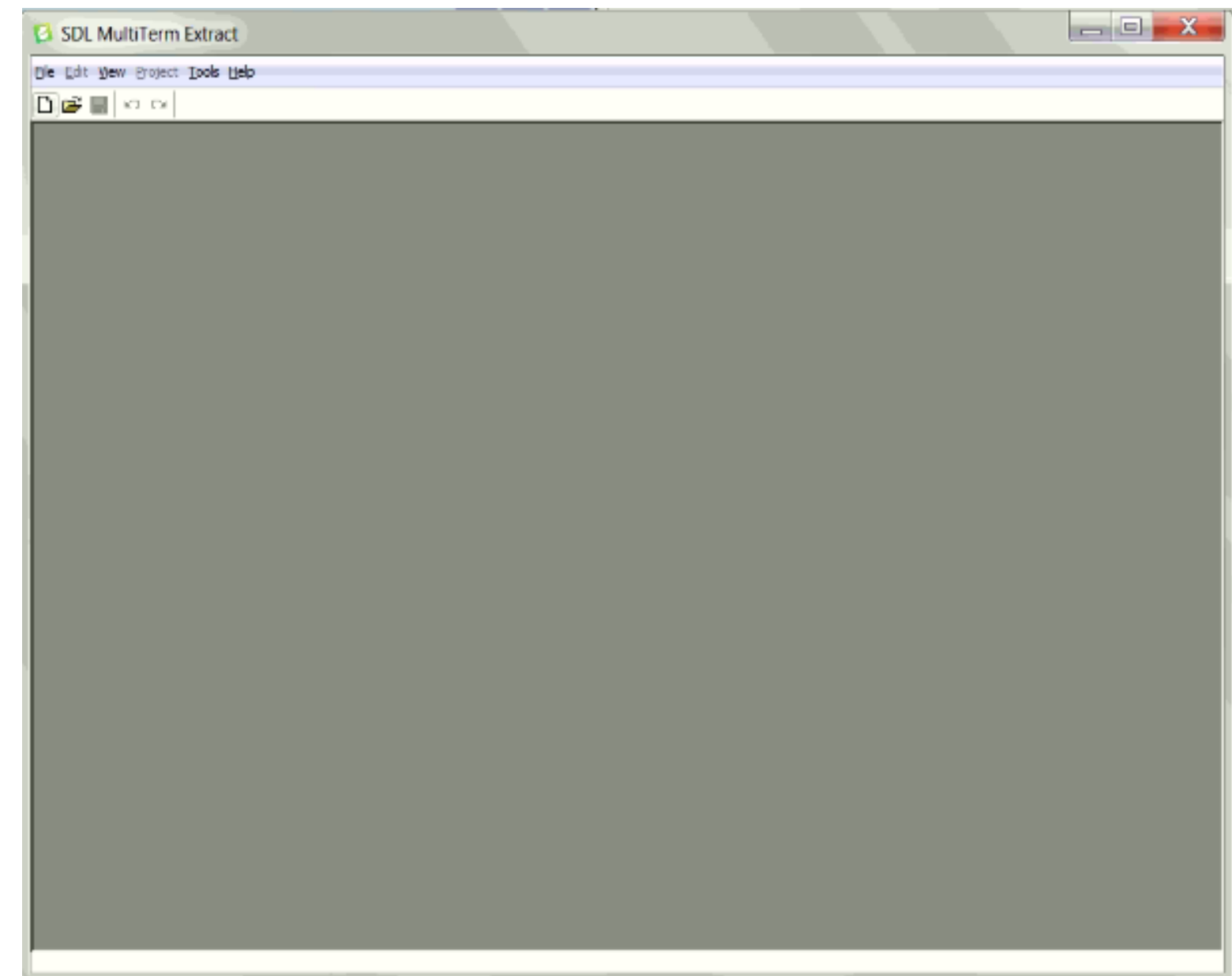
Utility software help convert Excel files into termbases (.sdltb) and vice versa as needed



C:\Program Files (x86)\SDL OpenExchange\Glossary Converter\GlossaryConverter.exe

SDL Multiterm Extract (EXTRACT)

Software to help identify and export suggested terms after analysing English files (monolingual mode) or translation memory files (bilingual mode)



C:\Program Files (x86)\SDL\SDL MultiTerm\MultiTerm14\MultiTerm Extract.exe

Key Technology Tools

Utilising other key tools on Steering

Dataset Cleanup & Sequencing

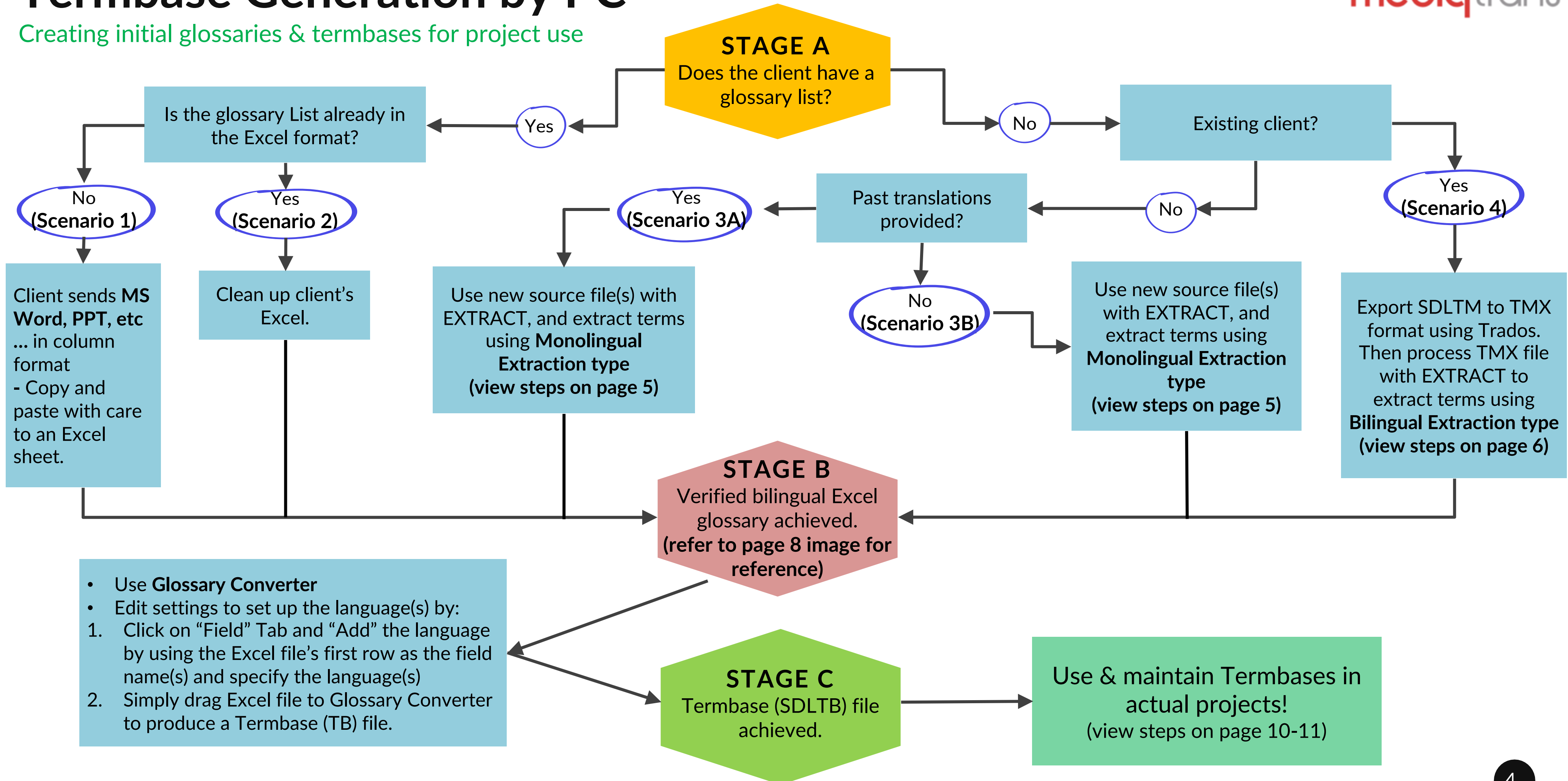
Utility function and process to help process and convert Extract-exported .txt file into cleaned Excel file

Client Glossary Combiner

Utility function to merge multiple XLSX glossary files together. This process automatically eliminates duplicates and also accumulates multiple target options per source term

Termbase Generation by PC

Creating initial glossaries & termbases for project use



Scenario 3A/3B – to reach stage B (using Monolingual Extraction)

Creating a human verified bilingual Excel glossary using available documents

Identify possibilities

- Create new EXTRACT project using “**Monolingual Term Extraction Project**”:
- Select **No Termbase** and specify the Source Language
- **Add files to the project** (.doc, .docx, .txt, .ppt, .pptx, .xls)
- Referring to page 7:
 - Specify exclusion lists via the **Excluded terms settings**
 - Specify desired terms via the **Term Extraction Settings**
- Note the number of terms (adjust silence/noise ratio and filter to achieve target quantity on page 8):



Export to .txt

- Export the identified terms to a tab-delimited text file:
 - In EXTRACT, click on **File** and select **Export**
 - Select **Create a new export definition** and choose **Tab delimited**
 - Specify the **Export file location**
 - Specify the **Filter** again (the one that was used earlier)
 - Click **Finish**

Clean up/finalise

- Curation by linguist to finalise the desired terms:
 - Utilise the Dataset Cleanup and Sequencing Tool (locate the tool on Steering; see page 3) and follow the steps to generate interim Excel
 - Send interim Excel to linguist
 - Instruct linguist to delete/omit any proposed terms that are incorrect/unimportant from the Excel
 - **Translate the terms:**
 - If **3A**, linguist will search for the appropriate translations using the client’s past translations
 - If **3B**, linguist will enter appropriate translation for each term
 - Linguist will save and send back a **verified bilingual Excel glossary**

Scenario 4 – to reach stage B (using Bilingual Extraction)

Creating a human verified bilingual Excel glossary using available TM

Identify possibilities

- Create new EXTRACT project using “**Bilingual Term Extraction Project**”:
 - Select **No Termbase** and specify the Source Language & Target Language
 - **Add files to the project (TMX)**
 - Referring to page 7:
 - Specify exclusion lists via the **Excluded terms settings**
 - Specify desired terms via the **Term Extraction Settings**
 - Specify the **Translation Settings**
 - Note the number of terms (adjust silence/noise ratio and filter to achieve target quantity on page 8):



Export to .txt

- Export the identified terms to a tab-delimited text file:
 - In EXTRACT, click on **File** and select **Export**
 - Select **Create a new export definition** and choose **Tab delimited**
 - Specify the **Export file location**
 - Specify the **Filter** again (the one that was used earlier)
 - Click **Finish**

Clean up/finalise

- Curation by linguist to finalise the desired terms:
 - Utilise the Dataset Cleanup and Sequencing Tool (locate the tool on Steering; see page 3) and follow the steps to generate interim Excel
 - Send the interim Excel to the linguist
 - Instruct linguist to delete/omit any proposed terms that are incorrect/unimportant from the Excel
 - Translate the terms by selecting the most appropriate translation as shown in the Excel
 - Linguist will save and send back a **verified bilingual Excel glossary**

Configuration Settings in Extract

Settings while using EXTRACT to identify possibilities (candidate terms)

Excluded Term settings

- Usage: Monolingual, Bilingual
- Use the following exclusion lists:
 - **MultiTerm Extract basic vocabulary lists** (already preloaded, don't need to do anything ... it contains name of countries, months, etc).
 - **List of fonts name** (e.g. Arial, Tahoma, etc)
 - Ensure the settings has included a "fonts_exclusion_list.txt" which can be found on desktop/WE-Angela/WINDA/TERMBASE/

Term Extraction settings

- Usage: Monolingual, Bilingual
- **Minimum and Maximum term length** - set the min. as 1 and max. as 5.
- **Maximum Number of Extracted terms** – leave unchecked.
- **Noise/Silence Ratio**
 - *Set ratio to maximum silence (and slowly decrease accordingly) to obtain target number of terms (refer to page no. 8 for target quantities).*
 - *When using maximum silence, if number of extracted terms are still more than the target quantity, use the filter (>80 or adjust level manually) to achieve the target amount.*

Translation settings

- Usage: Bilingual only
- **Maximum Number of Translations** – max. number of suggested translations for each term (set as 7).
- **Minimum Translation Frequency** – frequency of occurrence of term in the TM (set as 3).

Glossary Generation Strategy for Different Accounts Types

Budgeted hours and number of extracted terms.

			Before linguist verification	
Account	Budget	Final Outcome after linguist verification (glossary list)	Target quantity to export using EXTRACT	Target quantity after using Dataset Cleanup and Sequencing Tool
Tier 1 Account	3 hours	130-150 terms	230 terms	approx. 25% above Final Outcome
Tier 2 Account	2 hours	85-100 terms	160 terms	
Others	1 hour	40-50 terms	85 terms	



* For Glossary, we assumed translation productivity of 150 words/hour.

* Average number of words per term is 3. Therefore, term translation productivity is 50 terms/hour.

* Productivities above include time taken to perform term cleanup i.e. delete/omit incorrect/unimportant terms

Stage B Screenshot

What the bilingual glossary list should look like

English	Korean	Thai	Vietnamese
accurate position-orientation	정확한 위치-방향	ตำแหน่ง-ทิศทางที่แม่นยำ	định vị chính xác
adapter	어댑터	หน้าต่อ, อะแดปเตอร์	bộ phận chuyển đổi
adhesion	접착력	การยึดติด	độ bám dính
adhesive	접착	ยึดติด, ติด, กาว	...
adhesive system	접착 시스템	ระบบการยึดติด	...
Advanced Engineering	고급 공학	วิศวกรรมขั้นสูง	...
advanced force	개선된 힘	ระบบแรงขั้นสูง	...
Allen	육각	หกเหลี่ยม	...
aluminum fingertip	알루미늄 핑거팁	ปลายนิ้วอะลูมิเนียม	...
arbitrary	임의	ตามความพอใจ	bất kỳ
arbitrary orientation	임의의 방향	ทิศทางตามความพอใจ	hướng bất kỳ
Automatic Payload	자동 페이로드	คำนวณน้ำหนักบรรทุกอัตโนมัติ	Tải trọng tự động
autonomous piezoelectric cleaning system	자동 압전 세정 시스템	ระบบทำความสะอาดเพียโซอิเล็กทริกอัตโนมัติ	hệ thống vệ sinh áp điện độc lập
bio-inspired	생체 모방	แรงบันดาลใจจากสิ่งมีชีวิต	cảm hứng sinh học
bionics	생체공학	ชีวประดิษฐ์ศาสตร์	kỹ thuật sinh học
built-in	내장	ในตัว	tích hợp
cables	케이블	สายสัญญาณ, สาย	cáp
center pointing	중앙 포인팅	ชี้ศูนย์	định vị tâm
changeover	전환	เปลี่ยนสายการผลิต	thay đổi thiết bị
changer	교환 장치	อุปกรณ์เปลี่ยน	bộ thay đổi
channel	채널	ช่อง	kênh
circular movement	원형 움직임	การเคลื่อนที่เป็นวงกลม	chuyển động tròn

Note that first row of the Excel is to contain the the name of the language in each column

Using Termbases during Project Execution

Roles & responsibilities for project execution after Termbase has been created

Analyse (AM)

- Confirm availability of glossaries with client (not for each request but at required intervals). Send glossaries to PC team for processing if any
- Use available Termbase (SDLTB) together with client's source file(s) for quote (file analysis)

Red over-line in SDLXLIFF source segment, and data in Term Recognition

Execute (PC)

- If client sends additional/multiple glossaries over time, you can use the Glossary Combiner Tool (see page 3) to assist you
- Ensure the SDLTB is included in the hand-off to the linguist
- Instruct linguist to use the SDLTB while translating the SDLXLIFF
- Instruct linguist to enter new critical terms into SDLTB if any are observed during the translation process

Maintenance (PC)

- Post-translation, linguist is to submit SDLTB along with SDLXLIFF
- SDLTB is to be uploaded into Box for use in future project (refer to page 11 for details)

Ensure that new SDLTB file is larger in size than existing one

Termbase Naming & Storage

How and where to store termbases

Storage

- Create new folder in Box for each account as follows:
 \Box\Intranet\Production\PM\Live Projects\ [ENTITY]\[ACCOUNT FOLDER]\Translation Memory\Glossary

Naming convention

- Use “account.sdlb” for our client account (do not use novotech.sdlb; instead, always use **account.sdlb**)
- In the event that our client is an agency with different end-clients, each with different terminology preferences, as such we should save unique termbases using the name of the end client in small caps, to be stored in the same folder as account.sdlb
 e.g. oracle.sdlb, novartis.sdlb, equinix.sdlb

* For MediqTrans - there is normally only one Termbase for each client account (only exception might be medcomms agencies)

* For Attunemen - there might be more than one Termbase per client account, i.e. one Termbase for each end-client

Future Development

Areas of improvement for future

1. To allow Fast Quote (Steering automation) to automatically include Termbase file in each auto-quote
2. To add terminology check into current Xbench QA step