PharmaSUG China 2023 - Paper DS-165

An Automatic Annotation Tool to Facilitate NMPA ADaM package Submissions

Tingting Liu, ICON;
Jiaqi Song, Janssen R&D;
Jing Wang, Janssen R&D;
Guangjian Fu, Janssen R&D

ABSTRACT

According to the requirements of the “Guideline on the Submission of Clinical Trial Data (for trial implementation)”, Programming codes submitted in the submission package should be annotated with comments, understandable, executable, and not include external program calls, which in particular avoid using external macro programs. For many studies, the comments in programs by default are in English. To make submitted programs readable and understandable, comments need to be translated into Chinese annotations. Annotation translation usually takes a lot of resources and time, while different studies have a lot of repetitive work when preparing NMPA submission codes. Based on the requirements of the above guideline and the problems faced in work, an automatic tool was developed to help complete the annotations translation. The translation process involves comments extraction, matching with an established Chinese-English Comparison library, machine translation, review of the translation results, and retransmitting Chinese annotations back to code. With the use of an automatic annotation tool, the entire annotation translation process can approximately save two-thirds of the time and resources compared with legacy practice.

INTRODUCTION

To ensure that submitted programming codes are understandable and readable, and meet the requirements of the “Guideline on the Submission of Clinical Trial Data (for trial implementation)” ([1]), the English comments in programs need to be translated into Chinese annotations. The general practice was open program code files and translate the English comments one by one manually. It was time-consuming, tedious, and not efficient because the same/similar comments were translated multiple times within a study or even within a code file and further we have to do repetitive work for future studies.

Are you still troubled by this repetitive boring work? An automatic annotation tool had been developed to solve this pain point. This tool was developed based on SAS 9.4, including two SAS macros. Figure 1 shows the development approach of the entire process. It can be divided into four steps, development of a Chinese-English comparison library, comments extraction, translation and review of translated results, and retransmission of Chinese annotation back to code.

To facilitate the distinction, "comments" in this paper refer to English comments, and "annotation" refers to Chinese annotations.

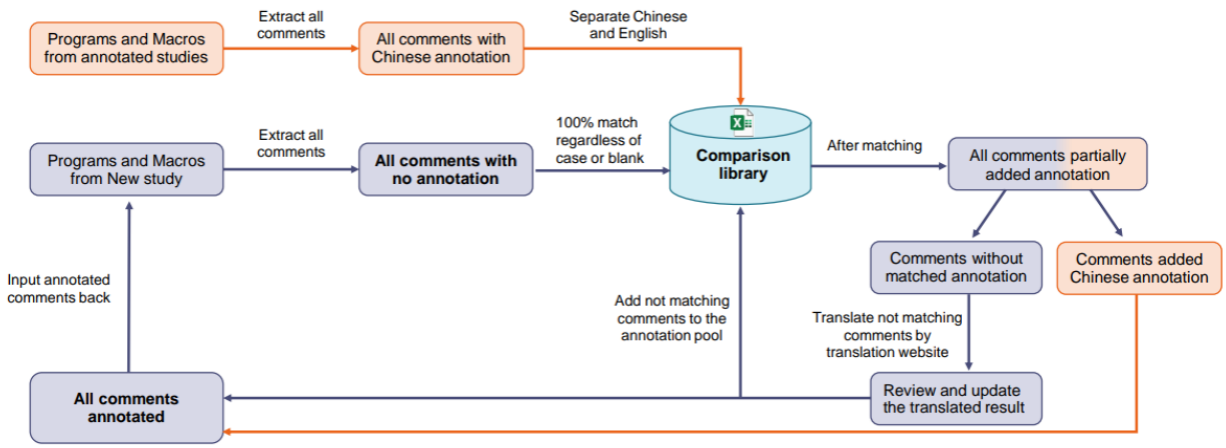

Figure 1 Translation process

INITIALIZATION OF THE CHINESE-ENGLISH COMPARISON LIBRARY

Firstly, initialize a Chinese-English comparison library is highly recommended because it not only greatly saves our time for English comments translation in future work but also reuses the translation results which have been completed before. The Chinese-English comparison library was developed from the previous submitted packages. All English comments and Chinese annotations extracted and exported to an excel file. As shown in below Example 1, the two columns in comparison library excel file were named as encom (the English comments) and cncom (the corresponding Chinese annotations).

encom	cncom
add attributes and output data set;	添加属性并输出数据集
clean library;	清理数据库
combine all the records	合并所有记录

Example 1 Chinese-English comparison library

COMMENTS EXTRACTION - %EXPOANNO

The second step is extracting English comments from the new study and automatically matching the English comments against with Chinese-English comparison library. The user needs to invoke the SAS macro %EXPOANNO to finish this step. There are 3 parameters in this macro as shown in Example 2.

libpath is the path that stores Chinese-English Comparison library Excel file.

prgpath is the path of programs that need to translate English comments. "|" is used as a separate symbol when there have multiple paths.

outpath is the path of the output file with partially English comments adding matched Chinese annotations and name as Export_Comments.xlsx.

The format of output Excel file as shown in Example 2. In addition to English comments and Chinese annotations (if they were matched in the library), there also contains some extra information that helps subsequent review, such as the program name(program), the position/record number of the annotation(recordn) and flag variable. Flag equal to ”Y” indicated this English comment was found in Chinese-English comparison library, and we did not need to review it later.

```
%ExpoAnno(libpath=,
          prgpath=,
          outpath=);
```

encom	cncom	program	recordn	flag
add attributes and output data set;	添加属性并输出数据集	adlb	688	Y
add brcafl variable		adsl	410	

Example 2 Export_Comments.xlsx

If no Chinese-English comparison library was created before, a blank Excel file with 2 columns (ENCOM and CNCOM) can be as the initial comparison library. After running this macro, you will get a file with cncom and flag columns empty.

MACHINE TRANSLATION AND REVIEW

For comments that do not have their annotation matched in the comparison library, they will be processed by the machine translation website or manual translation. The machine translation results need to be reviewed manually. After completing review, the Excel file that contained all Chinese annotations was created based on Export_comments.xlsx. Then you need to rename it as Import_Comments.xlsx (See example 3). This file will be used in the next step: retransmission of annotations.

encom	cncom	program	recordn	flag
add attributes and output data set;	添加属性并输出数据集	adlb	688	Y
add brcafl variable	添加变量 brcafl	adsl	410	

Example 3 Import_Comments.xlsx.

RETRANSMISSION OF ANNOTATIONS - %AUTOANNO

Another SAS macro %AUTOANNO had been developed to import Chinese annotations back into the code. Meanwhile, an extra excel file which including annotations that did not exist in comparison library were created to enrich the comparison library. The file will be appended to current comparison library and used in next delivering. All original programs were automatically backed up by creating a subfolder under program storage path. There have 3 parameters in %AUTOANNO.

libpath is the storage path of Import_Comments.xlsx.

prgpath is the path of programs that need to translate English comments. "|" is used as a separate symbol when there have multiple paths.

outpath is the path of output extra Chinese annotations Excel file.

Figure 2 showed how the program changes after processing. Figure 2a is the program before running %AUTOANNO and Figure 2b is the result after calling this macro.

```
%AutoAnno(libpath=,
          prgpath=,
          outpath=);
```

2a

```
options mprint mlogic symbolgen;
/*****************************************************************************
Merge the DM suppqual with parent domain by calling the %adjsupp macro.
*****************************************************************************/
%adjsupp(indsn=d_in.dm,
         insuppdsn=d_in.suppdm,
         inattrdsn=dm_in.valdef,
         memname=suppdm.qnam,
         outdsn=adsl,
         debug=OFF);

/*****************************************************************************
Convert rfidtc and dthdtc to numeric date variables.
*****************************************************************************/

%addtm(indsn=adsl,
       dtcvars=rficdtc dthdtc,
       adtcvars=,
       keepdtc=Y,
       outdsn=adsl,
       dropvars = ,
       debug = off);


/*****************************************************************************
  Keep needed variables
*****************************************************************************/
data adsl;
   set adsl;
    keep STUDYID USUBJID ARM ARMCD ACTARM ACTARMCD AGE AGEU SEX RACE RACEC COUNTRY INVNAM ETHNIC
        RFICDT DTHDT DTHFL DTHDTC SITEID SUBJID COHORT;
run;

/*****************************************************************************
Macros to add age and race groupings
*****************************************************************************/

%adagegr(indsn=adsl ,
         keepvars= agegr1 agegr1n agegr2 agegr2n ,
         outdsn= adsl,
         debug=OFF):
```

2b

```
options mprint mlogic symbolgen;
/*****************************************************************************
Merge the DM suppqual with parent domain by calling the %adjsupp macro.
通过调用%adjsupp宏将dm suppqual与父域合并。
*****************************************************************************/
%adjsupp(indsn=d_in.dm,
insuppdsn=d_in.suppdm,
inattrdsn=dm_in.valdef,
memname=suppdm.qnam,
outdsn=adsl,
debug=OFF);

/*****************************************************************************
Convert rfidtc and dthdtc to numeric date variables.
将rfidtc和dthdtc转换为数字日期变量。
*****************************************************************************/

%addtm(indsn=adsl,
dtcvars=rficdtc dthdtc,
adtcvars=,
keepdtc=Y,
outdsn=adsl,
           dropvars = ,
           debug = off);


/*****************************************************************************
Keep needed variables
保留所需变量
*****************************************************************************/
data adsl;
set adsl;
keep STUDYID USUBJID ARM ARMCD ACTARM ACTARMCD AGE AGEU SEX RACE RACEC COUNTRY INVNAM ETHNIC
RFICDT DTHDT DTHFL DTHDTC SITEID SUBJID COHORT;
run;

/*****************************************************************************
Macros to add age and race groupings
添加年龄和人种分组的宏
*****************************************************************************/

%adagegr(indsn=adsl ,
keepvars= agegr1 agegr1n agegr2 agegr2n ,
outdsn= adsl,
debug=OFF);
```

Figure 2 Changes in the programs after processing

Since Chinese annotations were added to the original programs, a quality check is recommended to ensure the new programing codes can reproduce the original results.

LIMITATION

Since the comments in SAS start with an asterisk and end with a semicolon or start with “/*” and end with “*/”, some normal code which contain special symbol may be mistakenly regarding as a comment and be wrongly extracted as shown below (Figure 3a). If there have commented out code, it will also be extracted as comments (Figure 3b).

If strange comments were found during annotation review, we can track and confirm them through the information in the Export_Coments.xlsx file (program name, record number).

3a

```
/*This is a test---There have star symbol in proc step*/
proc freq data = test1;
table usubjid*age;
run;
```

	A	B	C	D	E	F
	ENCOM	CNCOM	PROGRAM	RECORDN	FLAG	NOTE
	keep usubjid trt01pn trt01p age; sex race;		test	10		
	table usubjidage;		test	8		
	this is a test---there have comment out code in program;		test	7		
	this is a test---there have star symbol data or proc statment;		test	1		

3b

```
/*This is a test---There have comment out code in program*/
data test3;
    set adsl;
    keep usubjid trt01pn /*trt01p age*/ sex race;
run;
```

	A	B	C	D	E	F
	ENCOM	CNCOM	PROGRAM	RECORDN	FLAG	NOTE
	keep usubjid trt01pn trt01p age; sex race;		test	10		
	table usubjidage;		test	8		
	this is a test---there have comment out code in program;		test	7		
	this is a test---there have star symbol data or proc statment;		test	1		

Figure 3 Examples of wrong extraction

CONCLUSION

This automatic annotation tool enables the programming team to prepare code which meet the requirements of the “Guideline on the Submission of Clinical Trial Data (for trial implementation)” quickly, easily, and reliably. The annotation translation process can save approximately two-thirds of the time and resource compared with legacy practice.

REFERENCES

[1] “国家药监局药审中心关于发布《药物临床试验数据递交指导原则（试行）》的通告（2020 年第 16 号）” July 20, 2020. https://www.nmpa.gov.cn/yaopin/ypggtg/ypqtgg/20200720171201514.html.

ACKNOWLEDGMENTS

We would specifically like to acknowledge all members in Janssen NMPA e-sub initiative for their assistance and Baoqing Li, Peter Hu and Shechang Li for being our reviewers and offering great feedback.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Tingting Liu, ICON
Tliu82@its.jnj.com

Jiaqi Song, Janssen R&D
Jsong67@its.jnj.com

Jing Wang, Janssen R&D
Jwang493@its.jnj.com

Guangjian Fu, Janssen R&D
gfu5@its.jnj.com