

P1

Hi colleagues. My name is Loskutov Volodymyr. I'm a Statistical Programmer in Labcorp drug development. I and my team are developing different utilities automate and simplify different business processes. Today I want to tell you about Intext utility. This utility allows to reduce manual time-consuming coping tables during creating Clinical Study reports.

P2

What is Clinical study report? CSR is final step of any clinical trial before submitting the results to the regulatory agencies. It provides many details about methods and results of a medical trails. Usually this is a very long document with some tables and figures.

CSR is the result of cooperation of hundredths people, and you cannot fail with it.

For example: Clinical Study Report (CSR) for AstraZeneca trial 5077US/0049 has 70 tables, 10 figures, 8000 pages

P3

Which problems do we solve?

1. Updating Tables and Figures in report
2. Updating titles and footnotes
3. Convert sas7bdat dataset to tables

Let's talk a little bit more about each of them.

CSR (Clinical study report) may go through multiple rounds of recompiling during the development. You need to update all the tables and figures during each round. This is manual and time-consuming task when you need to copy some data from one to another document. Usually, it takes up to 5 minutes for each table.

Some tables also may have graphical issues (like a wrong pagination, lost titles or footnotes or data shifts). In this case, you must inform programming team and wait till they update and verify tables.

P4

What was the initial task?

We need to create the utility which may help to create to Clinical Study report.

Utility may insert tables and figures on special places inside the document.

It uses sas7bdat dataset to create table. Why? Because verifying sas7bdata dataset is faster and easier than verifying tables.

It uses jpeg images to create figures. Because figure is the image with a title and a footnote

Read titles and footnotes from the special file (for example Mock shell)

P5

Let look on our solution. We create windows applications base on Python 3.7. There are 5 different screens with simple graphical user interface. 4 screens to choose required files and tick some options and fifth screen with log window. Let's look on each window

P6

First Screen allows to select objects to insert.

Sas7bdat datasets used to create tables

Jpeg and png images used to create figures

P7

Second screen used to choose file with table numbers, titles and footnotes.

Look at the Mock shell as an example.

P8

Third screen used to choose CSR template. This is a text document with special tags (<Table 1>, <Figure 3>). Tags will be replaced on tables and figures.

P9-P10

Look on the results.

From the left side you can see a part of CSR template with tag and final CSR with table on the right side

One more example with figure.

P11

Make some Calculations. As an example, I will take CSR for AstraZeneca trial 5077US/0049

How much time we need to update it manually?

There are 70 tables and 10 figure. 5 minutes for each. 400 minutes or 6 and half hours total. Usually, you may find some cosmetic issues and you will wait for their fix (due to Quality control process). And 1 hour for manual check. Near 12 hours at all

Automated solution. 1 minute to run the utility. Utility spends up to 30 seconds (usually less) to process 1 object and insert it to CSR. 80 objects * 30 seconds = 40 minutes and 1 hour for manual check. Total time is less then 2 hours.

P12.

Summarize benefits of using automatized approach

1. Minimize time to update the report
2. Reduce chance of typo in titles and footnotes
3. Reduce chance of cosmetical issue in final tables and figures

P13

Programming background. There are 2 modules in Intext utility. GUI and doc_util

We used tkinter library for Graphical user interface. This is standard python library for simple interfaces

Doc_util module contains all business logic. There are 2 important libraries inside: docx and sas7bdat.

Sas7bdat is the library for Reading data from sas7bdat files and convert them to DataFrame object

Docx is the library for creating and updating Microsoft Word files. Docx.Oxml is a sub library or docx library. It used to Create tables in Microsoft Word files and specify their size, borders and fonts

P14

Our plans.

We are going to finish testing phase and try our utility on real project.

Right now, we compile all the files, but we are going to recompile only updated files. This must reduce CSR creation time.

We also going to add more client features. Like special rules for table border creation, title extraction etc.

P15

Right know my country suffers from Russian invasion. My friends from volunteer organization of paramedics "Hospitallers" need some medicine. As like us, they work for the sake every life. You can find fore info on their website <https://www.hospitallers.life/>

P16

Thanks for your attention

