


Decoding Program Creation Support Tool



Fumihiko Akazawa(JAMSTEC)

Motivation and Aims

- **JAMSTEC have deployed many types of floats so far;** Apex, ARVOR, Navis, NEMO, NINJA, NOVA, PROVOR, Apex-Deep, and Deep NINJA.
 - **The data format and parameter name are frequently modified by float manufacturers, even if float types and purchased year are the same.**
 - **Every time we have to work to find what are differences and where are corresponding parameters in comparison with previous ones. We waste a long time to identify and modify the decoding program.**
 - **Therefore, I have developed the tool to help us to find differences by comparing the data formats of the same type float which is new purchased float with previous one. It outputs a list of differences between the two, including alternative candidate parameter names in the decoded data by the Jaro-Winkler Distance method,** which can quantify similarity of character string.
-  **Using this tool, DACs and PIs can shorten the time to find parts of decoding program which should be modified and contributes to efficient decoding program creation.**

What does this tool do?

- This tool is developed by python.
- This tool is currently applicable to APEX and Navis.
- Input files for this tool are the following;
 - (a) msg and log files of the newly purchased float
 - (b) a list of parameters of the transmission data file of same type floats which were already launched (excel file).
- This tool compares (a) to (b), and output whether each parameter in (a) exists in (b) or not. Furthermore, for the parameter that does not exist, this tool searches for parameters that are close to the item names in (b), by using the Jaro-Winkler Distance method (*1).
- We will develop further tool to apply netcdf files by adding the information of netCDF file name (prof, traj, and tech) and parameter name in it corresponding to each parameter of the list (b). It is useful when creating netCDF file directly from the transmission file of floats.

(*1) Jaro-Winkler Distance method can quantify the similarity of character strings.

https://www.researchgate.net/publication/243772975_String_Comparator_Metrics_and_Enhanced_Decision_Rules_in_the_Fellegi-Sunter_Model_of_Record_Linkage

Samples and Future work

```
localhost : $ python3 afc.py [ xlsx ] [ newfloat ]
```

```
ActiveBallastAdjustments is probably AltDialCmd ( 66.0% )  
or AtDialCmd ( 63.0% )  
or CpActivationP ( 61.0% )  
BuoyancyPumpOnTime is probably BuoyancyNudge ( 95.0% )  
or BuoyancyNudgeInitial ( 94.0% )  
or ConnectTimeOut ( 60.0% )  
IceMLMedianT is probably PnPCycleLen ( 62.0% )  
or CpActivationP ( 55.000000000000001% )  
or AscentTimeOut ( 54.0% )  
.  
.
```

Users just type excel file name and transmission file name. [xlsx] is (b) and [newfloat] is (a) in the previous side.

List of candidate parameter names and probability.

- <https://github.com/argojamstec/ArgofloatChecker>
 - Now in under public review
 - Apache License Version 2.0
- Future version will also support NetCDF.