Fusion Engineering and Design 92 (2015) 1–7

Contents lists available at ScienceDirect

Fusion Engineering and Design

journal homepage: www.elsevier.com/locate/fusengdes

A GUI-based intuitive tool for analyzing formats and extracting contents of binary data in fusion research

O. Naito*

Japan Atomic Energy Agency, 801-1 Mukouyama, Naka, Ibaraki 311-0193, Japan

HIGHLIGHTS

- A GUI-based intuitive tool for data format analysis is presented.
- Data can be viewed in any data types specified by the user in real time.
- Analyzed formats are saved and reused as templates for other data of the same forms.
- Users can easily extract contents in any forms by writing a simple script file.
- The tool would be useful for exchanging data in collaborative fusion researches.

ARTICLE INFO

Article history:
Received 9 May 2014
Received in revised form
16 September 2014
Accepted 1 December 2014
Available online 9 February 2015

Keywords:
Binary data
Format analysis
Contents extraction
GUI-based tool
Collaborative fusion research
XML

ABSTRACT

An intuitive tool with graphical user interface (GUI) for analyzing formats and extracting contents of binary data in fusion research is presented. Users can examine structures of binary data at arbitrary addresses by selecting their type from a list of radio buttons in the data inspection window and checking their representations instantly on the computer screen. The result of analysis is saved in a file which contains the information such as name, data type, start address, and array size of the data. If the array size of some data depends on others that appear prior to the former and if the users specify their relation in the inspection window, the resultant file can also be used as a format template for the same series of data. By writing a simple script, the users can extract the contents of data either to a text or binary file in the format of their preference. As a real-life example, the tool is applied to the MHD equilibrium data at JT-60U, where poloidal flux data are extracted and converted to a format suitable for contour plotting in other data visualization program. The tool would be useful in collaborative fusion researches for exchanging relatively small-size data, which don't fit in well with the standard routine processes.

1. Introduction

In large scale scientific experiments such as those in nuclear fusion research, various kinds of diagnostic devices are involved and different kinds of formats are employed to save measured data. From the viewpoint of keeping the maintainability and extending the longevity of the data, adopting one of the widely-used standard formats for their storage would be an ideal way. In fact, in the field of nuclear fusion, the data format of MDSplus [1], which is a suite of software for data acquisition and data management, is used at many experimental devices. And in other scientific areas, formats for more general purposes like HDF5 [2] or netCDF [3] are used worldwide.

Although such standardization may be beneficial, for a diagnostic system that has been running for years, migrating its data to one of the modern standard formats may not be so easy. Apart from such device specific reasons, fusion research tends to be conducted more and more under international collaborations, because construction of new experimental machines requires lots of time and cost, and covering vast operational space by a single machine is not simple. In particular, in a comparative study of experimental results from different devices, data might be exchanged among collaborators in heterogeneous formats. To compare directly, they should be converted to the same format at some point. Many of the experimental devices have data conversion programs or wrapper routines to transform their own data to and from standard formats, so it is generally not a problem.

But there could be cases where the data do not fit in well with the standard routine processes. If the researchers want to obtain necessary information from the data quickly, they might need to write

* Tel.: +81 29 270 7326; fax: +81 29 270 7419.
E-mail address: naito.osamu@jaea.go.jp

http://dx.doi.org/10.1016/j.fusengdes.2014.12.039
0920-3796/

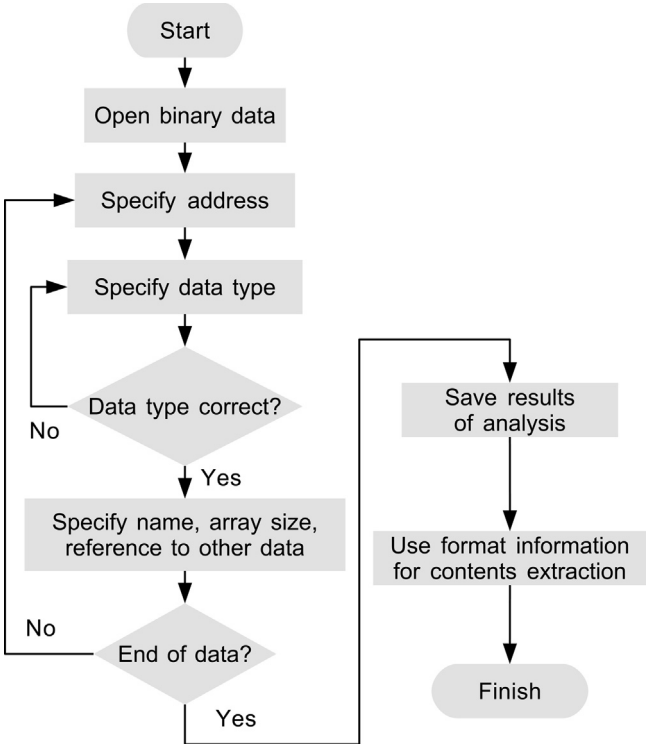

Fig. 1. Flow chart of data format analysis.

codes that read out the contents of data, but it is often cumbersome or time consuming. The objective of this paper is to alleviate such situation by developing an intuitive tool for analyzing formats and extracting contents of binary data in fusion research.

In the next section, requirements for such tool are explored. In Section 3, a simple model for specifying the data structure is presented. In Section 4, implementation of a proof-of-principle program with minimal features is described. In Section 5, a real-life application of this tool to fusion data is presented. In the last section, summary and discussion are given.

2. Requirements for the data format analyzing tool

First, we assume the users have at least minimal information to read the data, for instance, what kinds of contents are stored and where, although they may not be complete or the details might be missing. Otherwise reading arbitrary binary data would be virtually impossible. Based on that a priori information, the users would check whether the data at certain addresses were as expected or not. To do this, one of the popular methods is to display the data in a hexadecimal notation, and convert the portion of interest into the expected data type by hand, or using a scientific calculator, or writing a small data conversion program. If the result is reasonable, they record the format and repeat the process as needed. After finishing the analysis, they would compile the obtained information into a data accessing program and use it to analyze the data contents.

A flow chart for this data format analyzing process is shown in Fig. 1.

Based on this flow chart, the basic requirements for this intuitive GUI tool from the view point of the users would be as follows:

1 The users should be able to open any binary data file and view its contents in the hexadecimal notation on a computer screen.
2 On that screen, the users should be able to select any arbitrary byte to start data format analysis.

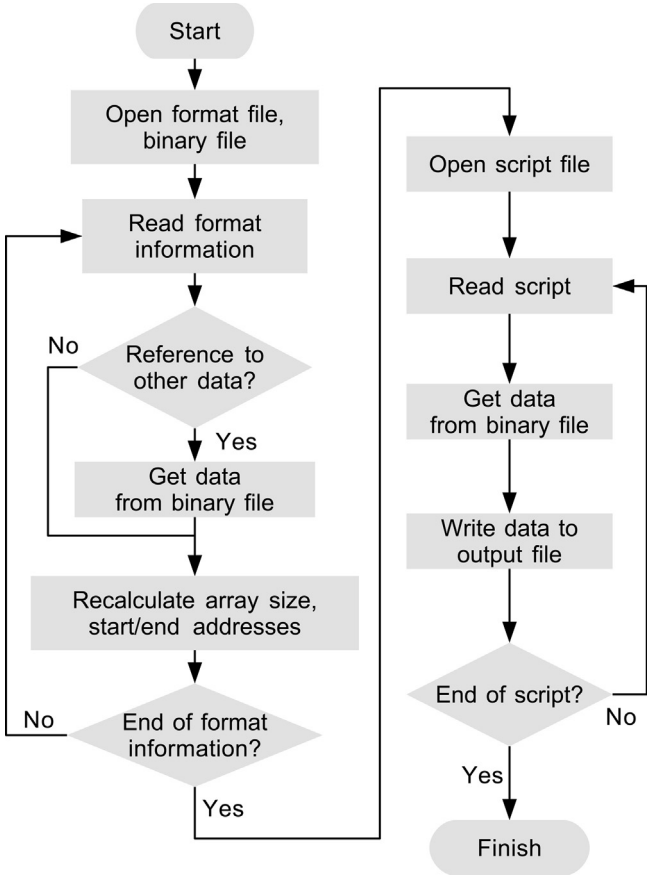

Fig. 2. Flow chart of data contents extraction.

3 The users should be able to view the data expressed in the data type of their choice in real time. Also, they should be able to specify the name and array size of the data. If the array size of the data depends on other data that appear prior to the former, the users should be able to specify the reference to the latter.
4 If the inferred data format seems reasonable, the users should be able to record that format. This information should preferably be hidden from the users so that they can use it without knowing its structures.
5 Based on the results of analysis, the users should be able to extract the necessary contents from the data with minimal effort. For that purpose, the users should not be required to write any complicated program nor know the contents of the format file.

With regard to the last requirement, one of the solutions would be to prepare a very simple script file for specifying what data are to be output in what format. A possible procedure to extract contents in the preferred format from the binary file, using the information in the data format file and the script file, is shown in Fig. 2. In this procedure, first the data format file and the binary file are opened, and the format information is stored into the memory. If the array size depends on the previously appeared data, they are retrieved from the binary file, and the array size and start/end addresses of all subsequent data are recalculated. Next the script file provided by the user is opened and the necessary data are fetched according to the recalculated addresses.

3. Description of the data structures

To specify data formats, a minimum schema for their description is necessary. To describe nested structures of data, two hierarchical