STUDY OF POSSIBILITIES OF DATABASE ANALYSIS OF TOKAMAK DATA USING ENCODERS [[1]](#footnote-1)\*)

DOI: 10.34854/ICPAF.2020.47.1.066

Kapralov V.G., Svintsov M.V., Bogdanov A.M., Novokhadskaya O.E., Totrov D.R.

SPbPU, Saint-Petersburg, RF, [v.kapralov@spbstu.ru](mailto:v.kapralov@spbstu.ru)

As the size of plasma systems and the cost of error increase, the importance of plasma control systems (PCS) increases [1]. For the stable operation of the tokamak, better forecasting of the evolution of the discharge is required, since the ratio of the energy stored in the plasma to the power of the active control systems increases and more time is required for their proactive switching on in order to return the plasma to an acceptable state. For the designed reactor facilities, such as TIN, the number of available diagnostic systems is reduced, compared with research facilities [2]. This is due to the fact that not all diagnostics can withstand neutron fluxes in a reactor, as well as with the transition from research to industrial tasks.

To set up and train control systems, tokamak data sets are needed, which are training examples for a typical situation from the point of view of physical processes in plasma or the technical task of automation [2]. As one of the options for analyzing the database and preparing training examples, the report suggests usage of existing powerful tools for working with text databases using SQL requests. The direct use of such tools is impossible or inefficient due to the storage of source data in the form of compressed binary sequences, accompanied by passports of signals and diagnostics. After extracting such a signal from the database, it is necessary to process it before deciding whether it refers to the case or not. This processing procedure significantly slows down the database analysis process.

To solve the problem, it is proposed to use an encoder to translate the source signals of tokamak diagnostics into the representation of text strings. As a result of a priori coding of the source database, a text database is formed, which is significantly smaller in terms of the required memory. Text database indexing allows one to quickly retrieve the number of shots or other information describing the shots with the desired event or physical phenomenon. Based on this data, if necessary, using decoding, it is possible to extract original tokamak diagnostics signals from the source database.

The report presents the development of an encoder, which includes the development of an alphabet for establishing correspondence between signal fragments and text symbols for their description. Based on the alphabet, a dictionary of combinations of symbols is formed that indicate a specific event or process in the plasma. And finally, using the dictionary, a text abstract is automatically compiled for the discharge as a whole. Set of these abstracts form the contents of a text database on the discharges of a tokamak.

The report examined the application of the encoder to bolometric diagnostic signals, including active circuits with a carrier frequency supplied to the sensor, and subsequent online processing of the response [3].

References

1. Dremin M.M. et al., Problems of Atomic Science and Tech., Ser. Th. Fusion, 2012 , **4**, 58.
2. Kapralov V.G. et al., Journal of Physics: Conf. Series, 2017, **907**, 1, 012027.
3. Kuteev B.V. et al., Review of Scientific Instruments, 2004, v.75, p. 4827.

1. \*) [abstracts of this report in Russian](http://www.fpl.gpi.ru/Zvenigorod/XLVII/Mu/ru/BW-Kapralov.docx) [↑](#footnote-ref-1)