Artificial intelligence in the problems of analysis and interpretation of omics human data

A. Shlikht*, N. Kramorenko
Far Eastern Federal University, Vladivostok, Russia
* e-mail: schliht@mail.ru

Key words: genomics, proteomics, common digital space, bioinformatics, artificial intelligence

Motivation and Aim: For efficient storage, analysis and interpretation of increasing omics data (BigData), it is necessary to convert it to SmallData using special types, formats, data structures [1]. The use of artificial intelligence technologies, databases, knowledge bases, expert systems are crucial here. The purpose of the work was to create an intelligent bioinformation system with effective structures and formats of presentation of omics data and knowledge of human.

Methods and Algorithms: The basis of the developed intelligent bioinformation system is the strategy of a common digital space (CDS). CDS is formed on the basis of highly structured, unified, complete and consistent registers, classifiers and coded dictionaries using standard structures and data types in the format of databases and knowledge bases. A characteristic feature of CDS is the presence of an autonomous information core, which focuses encoded highly structured data with a minimum of textual information, which allowed to move from BigData to Smalldata. The information core contains code parts of registers, classifiers and coded dictionaries (on the principle of machine – code, man – word). The creation of an autonomous information core has solved the problems of remote interactive access to world portals. This presentation of data in the core provides end-to-end information on all omics structures, many times increases the speed and validity of the analysis and interpretation of data. For filling of a CDS created by intelligent agents, which carry out conversion multi-format semi-structured data world portals, as well as the results of sequencing and mass spectrometry in a format the core CDS. This approach allowed moving from script procedural programming languages of 3rd generation (R, Phython) to the declarative languages of databases and knowledge bases of 4–5 generation, which frees the user from writing complex procedures for data extraction, analysis and interpretation. This presentation of data and knowledge makes it possible to build complex end-to-end logical chains in the tasks of genom-centered diagnosis and precision medicine.

Results: We developed an intelligent system of analysis and interpretation omics data, comprising: a stand-alone repository of highly structured data; a specific browser with a search system for navigating the space omics data and knowledge; subsystem knowledge extraction; subsystem access to world portals for updating data.

Conclusion: The created bioinformation system is an effective tool for biologists, physicians, geneticists, informatics, students in their scientific, practical and educational activities.

References
1. Shlikht A.G., Kramorenko N.V. (2016) Bioinformatic expert system of analysis and interpretation of omics sequence of the human genome. MM-HPC-BBB-2016:94.