A two layers hybrid approach for integration and first stage analysis of bioinformatics data based on hypergraph

Z. Sharifi[1], S. Kermanshahani[1], Z-S. shobbar[2]

[1]Computer engineering department, Faculty of engineering, Imam Khomeini international university, Qazvin, Iran

[2]Systems Biology Department, Agricultural Biotechnology Research Institute of Iran (ABRII), Karaj, Iran

Abstract: With rapid advancements in sequencing technologies, a large amount of data was generated in various data sources with the goal of accessing, analyzing and organizing them. To using of this valuable information for achieve a specific goal, we should integrate data from huge, heterogeneous, distributed data source into efficient way.[1], [2]. In this study, we provide an appropriate data integration approach for fetching and analyzing data. We apply this method to drought stress data in grains.Various approaches have been proposed for data integration. [1][4]. Each of these approaches follow different priorities. But, there may be different priorities for different data, one of the flexible methods that allows each of these approaches to be used according to their priorities is a hybrid approach.[5][6] Among all existing approaches to data integration, we have developed a special form of hybrid approach.[7].Biological information has complex structures, which means that they are unpredictable and have different relationships. These features make the data models such as graphs to be suitable for displaying and querying these structures. We can extract hidden relationships from this graph layer, and we conclude that the hypergraph is suitable for this second layer.[8], [9].In graph, relations are represented using the edges between entities. But in many biological processes, there are more than two entities, so these relationships can't be represented using the graph. Hypergraph provides a framework that can display such relationships without any limitations.[10]

Keywords: data integration; hybrid approach; graph data model; hypergraph; drought stress

References

[1] Z. Zhang, K. H. Cheung, and J. P. Townsend, "Bringing Web 2.0 to bioinformatics," *Brief. Bioinform.*, vol. 10, no. 1, pp. 1–10, 2008.

[2] Z. Zhang, V. B. Bajic, J. Yu, K. Cheung, and J. P. Townsend, "Data Integration in Bioinformatics : Current Efforts and Challenges," *Bioinforma. - Trends Methodol.*, no. 1, 2011.

[3] S. Hariharaputran, T. Töpel, B. Brockschmidt, and R. Hofestädt, "VINEdb: a data warehouse for integration and interactive exploration of life science data," *J. Integr. Bioinform.*, vol. 4, no. 3, p. 63, 2007.

[4] V. Lapatas, M. Stefanidakis, R. C. Jimenez, A. Via, and M. V. Schneider, "Data integration in biological research : an overview," pp. 1–16, 2015.

[5] N. Kausar *et al.*, *Informatics Engineering and Information Science*, vol. 253. 2011.

[6] R. Lathrop, "Heterogeneous Biomedical Database Integration Using a Hybrid Strategy," pp. 11–18, 2013.

[7] Z.-S. S. A. Bigdeli, S. Kermanshahani, "A Hybrid Bioinformatics Data Integration based on Graph Model," p. 2016, 2016.

[8] A. Birkland and G. Yona, "BIOZON : a hub of heterogeneous biological data," vol. 34, pp. 235–242, 2005.

[9] A. Lysenko, I. A. Roznov, M. Saqi, A. Mazein, C. J. Rawlings, and C. Auffray, "Representing and querying disease networks using graph databases," pp. 1–19, 2016.

[10] S. Klamt, U. Haus, and F. Theis, "Hypergraphs and Cellular Networks," vol. 5, no. 5, 2009.