

All purpose word pairing tool: Easy interaction networks for clinical data.

Thaynã Nhaara Oliveira Damasceno¹, Euzébio Guimaraes Barbosa¹

1 UFRN

Abstract

Considering that the hospital environment daily generates a large volume of data, and as well as the volume of published biomedical research, and therefore the underlying biomedical knowledge base, it's necessary to use specialized tools are able to transform data into information that influence in a positive way to decision-making in relation to clinical practice. The increasing of data available in the databases of organizational data, as well as the need of techniques that are most appropriate for its analysis has facilitated the emergence of new techniques for Data Mining, aiming at a better analysis of these. The DM can be defined as the process of discovery of patterns and relationships considered relevant within large data sets. As an extension of the Data Mining area, Text mining as being an application of computer systems involving both hardware and software in the textual analysis of documents. An algorithm called Integrate Paired Tool (IPT) was developed using the languages JavaScript, HTML, CSS, R, Perl and Shell Scripting. This tool enable quick tools to create interactive Gephi input files to plot Interaction Network from data described is lists. The algorithm is designed for a large variety of data, but it has a large impact to simplify data retrieved from clinical databases. The IPT uses techniques of Data Mining and Text Mining for analysis of Clinical Data, and these data can be aggregated by any professional in the multiprofessional team in health, not restricted to only one subarea. The tool has performed the analysis taking into consideration their own data supplied by the user. After pairing, the tool generates two files that can be displayed in the tool Gephi, one with the nodes and another with the edges of the network. Gephi is an open-source software for network visualization and analysis. Gephi allows the end user to operate, analysis, use of filters, the manipulation of data, as well as cluster and export of data from any type of networks. We hope that our tool could be further extended and used to analyze data from Pubmed queries in the future due to its powerful way to extract meaningful data from complex data files.

Funding: UFRN/CAPES/PROPESQ