

Text Mining Techniques for Prioritizing Pathogenic Mutations in Protein Families Known to Misfold or Aggregate

Authors : Khaleel Saleh Al-Rababah

Abstract : Amyloid fibril forming regions, which are known as protein aggregates, in sequences of some protein families are associated with a number of diseases known as amyloidosis. Mutations play a role in forming fibrils by accelerating the fibril formation process. In this paper we want to extract diseases that caused by those mutations as a result of the impact of the mutations on structural and functional properties of the aggregated protein. We propose a text mining system, to automatically extract mutations, diseases and relations between mutations and diseases. We presented an algorithm based on finite state to cluster mutations found in the same sentence as a sentence could contain different mutation cause different diseases. Also, we presented a co reference algorithm that enables cross-link sentences.

Keywords : amyloid, amyloidosis, co reference, protein, text mining

Conference Title : ICBBCB 2015 : International Conference on Bioinformatics, Biomedicine, Biotechnology and Computational Biology

Conference Location : Montreal, Canada

Conference Dates : May 11-12, 2015