BIOMEDICAL LITERATURE MINING SYSTEM FOR INFLUENZA A SUBTYPES


Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson Mississippi 39217, USA

Abstract: Over the past 100 years, there have been four influenza pandemics; H1N1 in 1918, H2N2 in 1957, H3N2 in 1962, and H1N1 in 2009 with each causing dramatic increases in death rates and hospitalizations although the pandemic of 2009 showed a decrease. The declaration in June 2009 of a pandemic of H1N1 influenza by the World Health Organization as well as past outbreaks of H5N1 influenza brings to the forefront the need to extract scientific information and knowledge on influenza subtypes that are buried in biomedical literature. The antigenic nomenclature of Influenza A viruses includes the combination of two groups of proteins: the Hemaglutinin (H) and Neuraminidase (N). There are 16 H types (H1-H16) and 9 N types (N1-N9) resulting in a total of 144 possible symbols. The Biomedical Literature Mining System for Influenza Subtypes was developed to facilitate information extraction from a collection of PubMed abstracts in which at least one of the 144 possible antigenic nomenclatures was mentioned. The resource available at http://compbio.jsums.edu/influenza currently holds over 7,700 PubMed citations. The collection can be searched for sentence level context of search terms. The resource is useful for selecting articles that are unique for a subtype or where co-occurrence of multiple subtypes exists. Since mutation is a mechanism for emergence of pandemic influenza viruses we have manually annotated the focus of a set of 933 sentences that contain the text “mutation.” Our long-term goal is to continue to develop the literature mining system to be useful to first responders in prevention, detection and recovery mechanisms of influenza epidemics or for another possible pandemic.

Acknowledgements: National Center for Biodefense Communication, Jackson State University; RCMI-Center for Environmental Health, Jackson State University; National Institutes of Health (NIH-NCRR G12RR13459; NIH-NIGMS T36GM095335; NIH-NIMHD 1P20MD002725-01; NIH-NCRR-P20RR016476; NIH-NCRR-P20RR016460); National Science Foundation (EPS-0903787; NSF-DBI-0958179; DBI-1062057); and US Department of Homeland Security Science and Technology Directorate (2007-ST-104-00007; 2009- ST-062-000014; 2009-ST-104-000021). Disclaimer: The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.