DISCOVERING PATTERNS OF EXPERIMENTAL FACTORS FROM BIOMEDICAL LITERATURE

Yogesh Pandit (Mathew J. Palakal),
School of Informatics, Indiana University Purdue University Indianapolis, Indianapolis, IN 46204

Abstract:
Background: Biological experiments are performed everywhere. Different biologists perform similar kinds of experiments with slightly different experimental conditions. These conditions are responsible for varied results. Question is; how do similar kinds of experiments give different results when the experimental conditions are changed? The experimental factors have to be playing an important role. The aim of our study is to explore experimental factors as a source of information to extract knowledge. Methods: Data is extracted from full-text articles. We have conceptualized the data representation as ‘bins’ which are filled by identifying conditions from the text. This is an exploratory study to discover patterns of experimental factors and how they vary with different experiments. We cluster documents to look for a thread of factors. The hypothesis is that there should be something in common and something that causes variation. We use clustering to put documents together based on statistical and semantic measures. Results: We clustered documents based on all the factors and compared these clusters with clusters from only a couple of factors. We observed the same documents were outliers in both. It means that out of all the factors, only a few cause the results to vary. As a sample study we used 75 documents in “lung cancer”. On comparing whole document clusters with Strain vs Gene clusters we found 3 documents as outliers. This sample study indicates that the combination of these strains and genes might be causing the conditional variation, and thus in the final results
too. Future work involves doing this exercise on a large set of documents and looking for the kinds of variations caused by these patterns of experimental factors.