Reverse PubMed Search Program

Sun, Mengyuan

The number of citations held in the PubMed database currently numbers more than 22 million. The reverse PubMed search program is a search and retrieval method that returns ideas for further research from search results instead of results from ideas. It can conduct a preliminary search for connections between two papers and return both the relationships and a table of values that helps determine each one's significance. The data collected from a run shows that this program can differentiate between relationships that are relevant to both research papers and those that are not. When the number of matches per word was greater than two, the word was likely to have a strong link to both subjects and hold a more significant connection. When the number of matches was two or less, the word was not likely to be related to science. When the frequency of unique words matched per two-paper dataset was high, the two papers were only similar. The program corrects for spelling and tenses of words and removes meaningless numerical text and stop words. This program was also able to find relationships that are potentially useful to students and researchers. Two papers on lung cancer and schizophrenia were related through cannabidiol because it was considered a treatment for schizophrenia and because it could inhibit lung cancer cell invasion. Researchers could use this information to investigate some of cannabidiol's biological and chemical mechanisms to explain why this relationship occurred or why it should not. The program still can be improved for better sensitivity and efficiency. When used in coordination with other search programs, it is another unique lens by which people can view and analyze data.

Awards Won:
Oracle Academy: Award of $5,000 for outstanding project in the systems software category.