

UTILIZATION OF A DATA MINING SOFTWARE FOR THE IDENTIFICATION OF EMERGING ANIMAL HEALTH ISSUES: BENEFITS AND LIMITATIONS

Bridges VE

Center for Emerging Issues, Centers for Epidemiology and Animal Health, United States Department of Agriculture, Fort Collins, Colorado, USA 80521

In today's information age, vast quantities of documents, articles, and other items containing information show up on one's desk either in hard copy or electronic form, much of it textual in nature. A major challenge is sifting through this information to determine what actually needs to be read in the limited time available to do so. To address this challenge, data mining software programs have become available that analyze large quantities of textual data, honing in on the files which are most relevant to a particular informational need.

The Center for Emerging Issues has been utilizing one such program (Pathfinder8¹) for over two years. In this presentation, our experience with the program is discussed, focusing on its application to emerging animal health issues. This evaluation of the software program assesses the benefits of its use in the identification and analysis of emerging animal diseases and other animal health issues, as well as including the program's limitations for this application.

The Pathfinder8 software makes use of any data which can be converted into electronic format. Electronic files are manually put into the database and hard copies of documents are scanned into electronic format and then added to the database. In addition, scripts were programmed to automatically receive information from list servers and to download information from specified sites on the World Wide Web. Complex queries consisting of multiple nodes, each with their own unique query, create a subset of the data. Analytical tools, which are a part of the program, are applied to this subset, allowing the analyst to quickly search, sort, cluster, map, and visualize the data. It also allows the analyst to perform link analyses which identify relationships between people, facilities, or other entities contained within the data.

¹ PRESEARCH Incorporated, Fairfax, Virginia USA