# Clustering Algorithm

The availability of news stories on the World Wide Web has increased dramatically in recent years. Although search engine algorithms have made it simpler to find these papers, users must still put forth a lot of work because the majority of search engine algorithms only scan for keywords and do not analyse the context of the entire article. We suggest a mechanism that groups articles into subjects. More particular, we have concentrated on using text mining techniques to assist in resolving issues presented by a media outlet or public relations division.

Data mining techniques are being used to address the problem of locating documents that may be of potential importance. More specifically, activities like finding trends in text documents have been accomplished using a specialized form of data mining called text mining. is a method, for instance, for figuring out subjects that appear repeatedly in papers. It finds similar groups of items by representing an article as a collection of entities. The solution we suggest, however, is significantly less comprehensive and more useful because it just seeks to address the issue of grouping articles with related topics. News organizations would prefer to have easy access to relevant documents. Current search algorithms place an excessive amount of weight on document popularity and other variables.

By providing information to the user, the method we propose greatly simplifies this task. Instead, then relying on specific keywords or popularity to determine search results, our method groups related articles together based on the overall content of each one. The automated clustering of relevant news articles is done by our text mining engine. The article clustering system we created is described in this paper. It is designed to operate off of a database of news organization-written stories. Unless the articles are already constrained to a specific scope of themes, a relatively high number of articles are required for this approach to yield useful results.