



**CALIFORNIA STATE SCIENCE FAIR  
2011 PROJECT SUMMARY**

<b>Name(s)</b> <b>Rahul Sridhar</b>	<b>Project Number</b> <b>S1426</b>
<b>Project Title</b> <b>Multi-document Summarization using Spectral Clustering</b>	
<b>Abstract</b> <b>Objectives/Goals</b> It is not uncommon in today's electronically connected world to get information about the same subject from a variety of sources. Search engines like Google have made this possible with a mouse click and as a result human beings are inundated with information. The challenge is to combine these "results" into a concise summary. Can this summary be automatically generated based on quantitative scores with no qualitative judgment? In other words, can we write a program that will create a concise, effective, and coherent summarization of multiple articles on the same subject? <b>Methods/Materials</b> My method uses the technique of spectral clustering to summarize multiple documents. Clustering techniques are widely used for data analysis and spectral clustering often outperforms traditional clustering algorithms such as k-means. Given a set of documents, the program will first create a similarity graph, with vertices representing the sentences in the documents and weighted edges between vertices to represent sentence similarity. Next, the graph will be divided into a certain number of clusters, where each cluster represents a group of sentences that are similar to each other. A representative sentence is then chosen from each cluster. These sentences are then ordered to create a summary. For my experiments, I chose news articles and results of search-engine queries as multi-documents. <b>Results</b> The proposed method is fast and effective for documents containing news articles and reviews. In addition, the results satisfy two key properties: (i) summary does not contain redundant information; (ii) sentences conveying little or no information are not included in the summary. <b>Conclusions/Discussion</b> In this reasearch, I implemented a method to generate a concise and accurate summary of multiple documents on a common subject. This research would have applications in many varied fields, for example in summarizing news articles or search results of specific topics. For future work, I would like to extend this technique to implement an "aggregator" that can collect multiple related news articles from a set of sources that feeds into the "summarizer".	
<b>Summary Statement</b> Given multiple documents on a related subject, automatically create a summary that is both coherent and accurate.	
<b>Help Received</b> My Mother helped me with proof reading my slides.	