



**CALIFORNIA STATE SCIENCE FAIR
2003 PROJECT SUMMARY**

Name(s) Jamie L. Mountford	Project Number S1220
Project Title The Ability of Optical Character Recognition to Accurately Recognize and Identify Word-Processed Characters	
Abstract Objectives/Goals To what extent can Optical Character Recognition accurately interpret individual word-processed characters? The objective is to determine if the characteristics of specific symbols on a document page affect the accuracy of an OCR output reading from that page. Methods/Materials Materials: computer, scanner, printer, Jet Suite Optical Character Recognition Software Procedures: A. Sample template was created on a word processor that included a variety of characters, including letters, symbols, numbers, and punctuation marks. B. From this template, 13 document pages were formatted, each with a separate typeface characteristic. C. Each of the separate pages was run through an OCR system and checked for accuracy. D. Accuracy was measured for each individual character based on the number of errors produced by the OCR system from the original document to the image file on the computer. Results Alphabet and numerical characters produced the most accurate OCR output overall. Foreign characters and symbols had lower accuracy percentages and a higher number of average errors. Conclusions/Discussion The uniformity and familiarity of the numbers and letters aided the software in producing accurate output documents. Because of the lack of programmed memory and lexicon capabilities, the unusual symbols were generally misinterpreted and produced inaccurate output documents.	
Summary Statement My project analyzed Optical Character Recognition output documents to determine what qualities of individual characters affect the accuracy of the system identification abilities.	
Help Received	