timesunion.com

Scientists get computers to grade handwritten essays

By CAROLYN THOMPSON, Associated Press
Last updated: 5:53 p.m., Thursday, January 17, 2008

BUFFALO, N.Y. -- With No. 2 pencils in hand, grade school students this month are scrawling essays in answer booklets to prove their language arts skills.

Then will come the educators' turn to be tested -- grading all those handwritten responses.

Right now, that means an army of scorers reading each word (itself a challenge given penmanship's falling cachet), and checking for reading comprehension, sentence structure and the like.

But researchers at the University at Buffalo say they're on to a better way -- software that allows a computer to do the work. Instantly.

Automated essay scoring that uses artificial intelligence to mimic the human grading process is not new on electronically written responses. It's standard fare, for example, during online practice runs for college prep tests.

And handwriting recognition programs are routinely used by the post office to read addresses off envelopes in sorting mail.

But putting the two technologies together to score essays handwritten by children was uncharted territory, said Sargur Srihari, a computer scientist and principal investigator on the project.

It could prove particularly useful in scoring standarized tests like the English Language Arts exam New York students are taking this month. No one expects the tests to be given electronically anytime soon, given the expense, necessary keyboarding skills and potential for network trouble, the researchers point out.

In preliminary testing, Srihari's team's computer program graded sample essays within one point of the way human graders did 70 percent of the time. The results will appear in the February/March issue of the journal Artificial Intelligence.

"We were able to do fairly well," said Srihari, who worked with the post office on its handwriting recognition capabilities. "It was our first foray into what can be done with children's handwriting."

A decade ago, James Collins, Srihari's collaborator from UB's Department of Learning and Instruction, would have said it couldn't be done.

"Now I'm absolutely convinced that computers can do many of the diagnoses that human readers do and respond in a similar fashion," he said.

Funded by a $100,000 National Science Foundation grant, the project used handwritten essays from Buffalo eighth-graders who were asked to read a passage and answer: "How was Martha Washington's role as First Lady different from that of Eleanor Roosevelt?"

The first step was scanning the essay into a computer and, using word recognition, turning it into a digital text file.

For the scoring step, the researchers taught the computer what key words, phrases and other features to look for and what values to assign, based on grades given by humans.

"The (question) says Martha Washington and Eleanor Roosevelt, so that is the vocabulary they're going to be using," explained Srihari, who is director of UB's Center of Excellence in Document Analysis and Recognition (CEDAR). "And there are other phrases and words we can get from the passage. It says, 'Martha Washington was hostess of the nation,' so there will be words like that ... So it is not a totally unconstrained problem where we are reading something out of the blue and the vocabulary was unlimited."

The benefits of the software go beyond saving time and money, Collins said.

"Right now, teachers spend a lot of time getting their students ready for these standardized tests, then the students take the exam and get their scores back months later," he said. "With computer scoring, students could get back their scores much faster at a time when the results can still be addressed. The assessment scores wouldn't just be going into a black hole."

Collins also envisions the software as a teaching tool that could be used to identify patterns or trouble spots in a child's writing.