"Searching With Comprehension: Going Beyond Google"

Patricia M. Brent, Nick Grabenstein and Tarik Umar, Oak Ridge High School, Oak Ridge, TN – 2005-06 National Team Finalists

Abstract: Natural Language Processing (NLP) deals with the extremely difficult task of allowing computers to understand, generate, and otherwise process human languages. One of the most prevalent NLP tasks is known as Named Entity Recognition (NER). Essentially, NER attempts to identify the categories of meaning to which certain noun phrases belong. Such classification of meaning through the use of lexicons alone is extremely error-prone. Classification via NER and other methodologies that build upon it has many applications. For example, finely structured databases are being designed in such diverse fields as biomedicine, national security, and business to allow efficient discovery of useful information. However, these specialized databases cannot incorporate information from unstructured text sources until the relevant information has been identified using NLP techniques. In this paper, we discuss the progress and future goals of our research applying statistical machine learning in NER. We discuss our GUI to speed up the tagging of text and our use of a Conditional Random Field (CRF) to identify standard NER types using both individual and combined models. In the case of individual models, we discuss our attempt to design a versatile and intelligent conflict resolver to boost efficiency and utilize overlapping information to improve performance.

"LZAC Lossless Data Compression: A Novel Approach to Minimum Redundancy Coding"

Allan Chu, Saratoga High School, Saratoga, CA – 2002-03 National Individual Finalist

Abstract: This paper presents a new dictionary-based universal lossless data compression algorithm, LZAC, that can be used to ease Internet traffic congestion, increase the effective rate of data transmission and increase the effective data storage in handheld and wireless devices. LZAC introduces two new concepts that have been mathematically proven to be optimal. LZAC runs in any buffer size.

Its performance was evaluated in 4k, 8k, 16k, and 32k-byte buffers using the Calgary Corpus benchmark. The results demonstrate that LZAC outperforms the major LZ77 variants in the 4k to 16k-byte buffers, as well as the fast version of GZIP in the 32k-byte buffer without resorting to statistics-based techniques. Thus LZAC retains the key LZ77 characteristics. LZAC is simple, fast in decoding, and economical in terms of memory consumption. These characteristics make LZAC especially suitable for the Internet as well as handheld and wireless devices. **Mentor:** Dr. Ken Chu

"Reality II: Creating a Fractal-Based Virtual Reality with Mass Appeal"

Matt Mousseau and Elliott Prechter, Lakeview Academy, Gainesville, GA -- 2001-02 National Team Finalists

Abstract: The purpose of this paper is to describe a product that we have been developing in which fractals are so clearly integral to the design that its distribution will spark interest in the field. The product will be a simulation program called Reality II, in which each user plays the role of sovereign over his territory upon a procedurally generated world of nearly infinite size.

The project is incorporating fractals in the following areas: a realistic economics simulator that models market fluctuations, a world generation algorithm that supports unlimited levels of detail based upon a modified fractal subdivision technique, and a world rendering system that incorporates fractals for visual effects.