NSCEE
National Supercomputing Center for Energy and the Environment
High-Performance Computing and Communications in Nevada

August, 2004
Vol. 14, No. 2

TeraWord

The Newsletter of High-Performance Computing and Communications in Nevada

Contents

NSCEE Collaborates with Health Level Seven, Inc. (HL7) to Develop Electronic Health Record (EHR) Standards

In April of 2003, the U.S. Department of Health and Human Services (DHHS) formally requested that Health Level Seven, Inc. (HL7) develop the standard for a national electronic health record (EHR), a welcome move on the part of DHHS that will go a long way in achieving the dual goals of controlling healthcare costs and improving patient outcomes.

The research staff of the UNLV HHS Telehealth and DOE Medical Records programs are currently members of the HL7 consortium. This consortium guides architecture and design decisions for the development of Version 3 of the HL7 electronic record system standard.

Recently, the UNLV DOE Medical Records team began supporting an initiative to re-engineer HL7's suite of software tools. This effort will be hosted on NSCEE's new Hewlett-Packard rx2600 Itanium 2 (Intel) Integrity Cluster with HL7 users located world-wide. These tools support the development of HL7's interoperability standards, which ultimately effect the success of implementing a standards-based EHR.

Founded in 1987, HL7 (http://www.hl7.org) is a not-for-profit, ANSI-accredited, standards developing organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. Level Seven refers to the highest level of International Standards Organizations (ISO) communications model for Open Systems Interconnection (OSI)-the application level. The seventh level supports such functions as security checks, participant identification, availability checks, exchange mechanism negotiations and, most importantly, data exchange structuring.

Why HL7? There are several health care standards development efforts currently underway throughout the world. HL7 is singular as it focuses on the interface requirements of the entire health care organization, while most other efforts focus on the requirements of a particular department. Moreover, on an ongoing basis, HL7 develops a set of protocols on the fastest possible track that is both responsive and responsible to its members. The group addresses the unique requirements of already installed hospital and departmental systems, some of which use mature technologies.

HL7 efforts enable effective, efficient communication between the constituents of the healthcare community which consists of an international community of healthcare organizations, vendors, healthcare information systems developers, consultants, systems integrators, and related public and private health services agencies. HL7's 2,000 members represent over 500 corporate members, including 90 percent of the largest information systems vendors serving healthcare.

Computational Consultant Joins NSCEE Staff

Jaime E. Combariza, Ph.D., recently joined the NSCEE staff as computational consultant. He received a B.S. in Chemistry and Biology from UPTC, Colombia, South America. He came to the U.S. to pursue graduate studies, obtaining a Ph.D. in Chemistry and Biochemistry from Southern Illinois University with a concentration in Theoretical and Computational Chemistry.

Jaime joins NSCEE from the University of Arizona where he worked in the Research Computing Support group, most recently as manager of the group.

Combariza brings a wealth of experience in High Performance Computing, programming skills, parallel programming, third party applications support, code optimization, local environment optimization and support, cluster and grid computing, to the NSCEE and UNLV research community. Combariza is an active researcher as well and has published over 20 articles in refereed journals.

Combariza's responsibilities include:

- Supporting researchers on the effective usage of computer resources.
- Assisting users on programming issues like code porting, code optimization, parallel programming, code debugging.
- Providing training workshops on the proper utilization of computer resources, new technology and programming techniques.
- Assisting users on data analysis using visualization tools.
- Working with research groups involved in collaborative projects with other institutions that will make use of Internet2 resources.

Research Activities at NSCEE

UNLV's Robot and Computer Vision Laboratory

The Robot and Computer Vision Laboratory (RCVL) was created as part of the Center for Cybermedia Research at the University of Nevada Las Vegas. Evangelos Yfantis, Laboratory Director and Professor in the UNLV School of Computer Science, founded the laboratory as an offshoot of his long-established Image Processing and Computer Graphics Laboratory at UNLV. In addition to Evangelos Yfantis, the laboratory employs two full time researcher scientists (Jerry Derby and Dimitri Papaioannou), and a host of graduate research assistants and student programmers. Dr. Yfantis brings more than thirty years of experience in digital signal processing to these laboratories. He and the research groups in these two laboratories have, under his guidance, generated an impressive record of publications in the field.

The RCVL currently receives funding from several Federal agencies and expects funding sources to increase as the laboratory continues to grow. The primary goal of the laboratory is to design new algorithms to perform computer vision tasks. In addition, the laboratory's team of programmers produces custom software to exploit new and existing computer vision techniques.

What is computer vision? At the most basic level, computer vision is a field of computer science that combines elements of artificial intelligence and digital image processing to analyze visual data. The field has strong ties to robotics since robotic sensory input must be converted to a digital input signal for processing by a robot's computer-driven decision making algorithms. The field is of growing importance as the cost of human decision making abilities continues to rise with respect to computer processing costs. Computer vision typically combines image processing technology to filter the input images with post-processing to perform pattern recognition. The identification results from the computer vision algorithms can be used to make decisions based upon the analysis of the input images.

One project currently under development at RCVL is part of the federally sponsored e-Health Records Project. This project is designed around a vast library of health-related records collected by the funding agency over more than five decades. Currently, the search for records can take months as humans must manually scan paper records to find those that match a given search criteria. The time, labor, and cost of these searches make all but the most essential searches problematic. In addition, the recent implementation of the Privacy Rule within the Health Insurance Portability and Accountability Act of 1996 (HIPAA) places additional restrictions upon human access to sensitive data within the stored medical records.

The goal of the RCVL is to produce software to analyze digitally scanned documents from the record library. Once the data is analyzed, then searches can readily identify pertinent records and requesters may receive digital copies of the records with virtually no time delay.

Recognition of textual data within a scanned document using the optical characteristics of the image has been actively pursued for several decades. There are numerous solutions (of varying accuracy) to problems of text recognition. However, this project introduces numerous complexities that cannot be solved using any existing software and requires development of new algorithms and the synthesis of numerous image processing and computer vision areas to reach a solution.

Some of the challenges of the project are briefly described below. The quality of some of the paper records has greatly deteriorated over time; advanced pre-processing algorithms are needed to enhance the image quality in order to optimize recognition rates. The most important data on the records is generally unconstrained handwritten text; this is the most challenging type of character recognition and solutions in this area remain weak.

Just identifying characters is insufficient for searching. The identified text must be linked to a label for searching. For example, the text "Doe, John" is of limited value in a search. Additional clues must be utilized to determine that "Doe" represents the surname of the patient and "John" is the first name for a more robust searching scheme.

One solution to identifying particular fields within a document is to use a filter to identify a form before parsing for text. Once a scanned page of text is identified, then the software developed by the laboratory can use stored information about the form to determine the location of particular data entered into a form. The RCVL has created a trial database of forms (a subset of the total set of forms to be used in the project) storing key identification information about these forms. Utilizing pattern recognition techniques, the software can identify which form was input with more than 99% accuracy.

One of the first steps necessary to identify text data is the breakdown of a scanned image into textual components; namely words and letters. In an effort to speed processing of documents, the RCVL has designed and implemented multi-phase partitioning algorithms. A fast algorithm will be used to perform the initial character separation task. The results of this separation are analyzed and then the program will look for atypical results. These atypical results (such as characters that are much wider than average - most likely representing multiple characters that ran together) will be re-analyzed using higher-cost algorithms to partition characters that were misidentified in the initial pass.

Another of the laboratory's more impressive accomplishments in computer vision is the recent attainment of 95% accuracy when recognizing unconstrained handwritten numeric data. The principal search field for records is the patient identification number or social security number. So this high accuracy in numeric recognition is a significant achievement towards successfully completing the project. A refinement of the numeric recognition should yield higher accuracy. In addition, it will be possible to integrate a database of known patient identification numbers with the decision making algorithms. Such information can be used as a final determinant in cases of marginal identification and provide even higher recognition rates.

The Robot and Computer Vision Laboratory at UNLV develops state-of-the-art solutions to problems in computer vision. For more information about the capabilities of the laboratory, contact Evangelos Yfantis at the UNLV School of Computer Science.

Research Activities at NSCEE

Logomarker and Digital Watermarking for HIPAA Compliance

Logomarker partially fulfills the requirements of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the subsequent standards adopted by the Secretary of Health and Human Services*. Logomarker provides assurance that electronic health records provided by the parent system (the UNLV's Medical Records Knowledge and Information Management for Radiation Workers program) are indeed uncorrupted electronic copies of the original documents. Logomarker achieves this with two technologies: visible watermarking and encryption.

Logomarker first embeds a visible Department of Energy logo into the electronic health record as a watermark. This logo is difficult to remove without damaging the original document, and it is also difficult to alter the information on the document without disturbing the logo. Thus the embedded logo visibly identifies the electronic health record as originating from the parent system, and also deters modification of the document. The image is then converted to a transmittable format (JPG compression).

Second, Logomarker embeds encrypted provenance data (information about the original document's source, author, date of conversion to electronic format, and a validation key string, among other information) into the header of the transmittable image. This does not affect the image in any way. The data is encrypted using standard encryption techniques that are

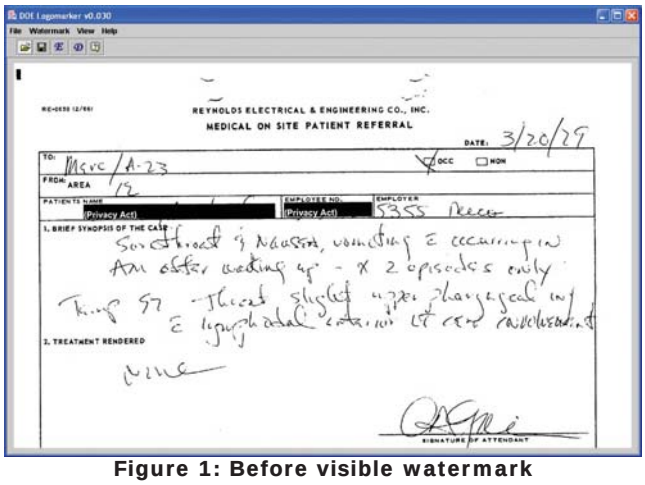

Figure 1: Before visible watermark

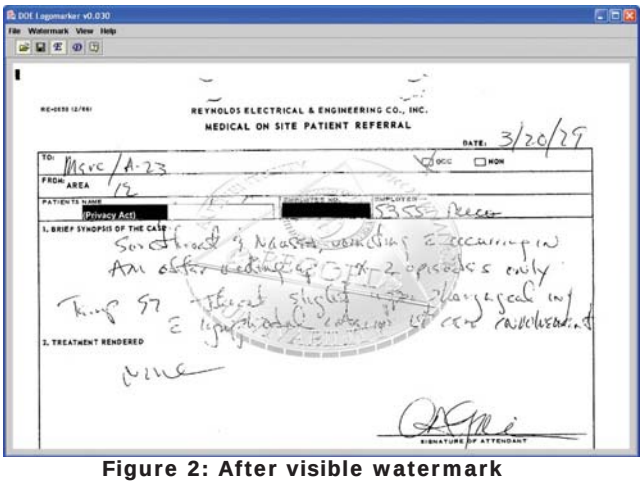

Figure 2: After visible watermark

difficult and time consuming to crack (3DES or AES), providing assurance that the provenance data has not been altered. The provenance data can be used to verify that the electronic health record originated from the parent system and that it has not been altered in transit.

Logomarker also has the capability of inserting transaction data into the header of the transmittable electronic health record. Transaction data includes the request time and date, the requesting user and computer, the source computer, and other data that uniquely identifies the record by time, location, and user. That is, even if the same user requests the same record from the same computer, the transaction "fingerprint" will be different because the time will have changed. This transaction "fingerprint" is another way to ensure that the electronic medical record is not altered in transit or by the user.

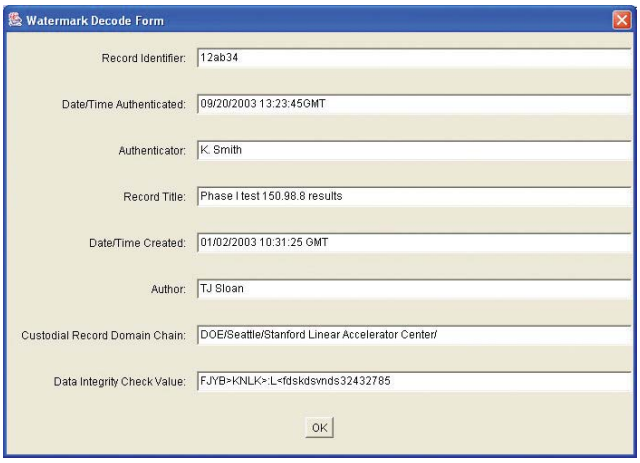

Figure 3: Decoded provenance data

Logomarker is implemented as a Java class. Normally, Logomarker would run "in the background", meaning there are no user-interactive parts to Logomarker. In order to demonstrate Logomarker's capabilities, a graphical user interface was developed to allow a user to select a document, embed the visible watermark and some generic test data, and extract the data at a later time. The figures in this document are taken from the demo version of Logomarker's graphical user interface.

Logomarker is being developed by the Center for Cybermedia Research, a consortia of semi-autonomous laboratories that focuses on the intersection of two well-established sub-disciplines of computing: computer networks and digital media. The laboratories, each with its own research agenda, that form the Center for Cybermedia Research are the Cybermedia Research Lab, the Image Processing and Computer Graphics Lab, the Internet Forensics Lab, and the Robot and Computer Vision Lab. For additional information, visit their web site at www.ccr.i2.nscee.edu.

* For related applicable requirements and standards, refer to Standards for Information Transactions and Data Elements, section **1173, (a)(1) and (d)(2)** of the Health Insurance Portability and Accountability Act of 1996 and Code of Federal Regulation (CFR) Title 45, Subtitle A -- Department of Health and Human Services, Subchapter C -- Administrative Data Standards and Related Requirements, Part 164 -- Security and Privacy, Subpart C--Security Standards for the Protection of Electronic Protected Health Information, **164.312(a)(2) and 164.312(e)(2)**.

NSCEE to Co-Sponsor International Conferences

ISNG (Nov 10-12, 2004, Las Vegas)

The **International Conference on Information Systems - New Generations (ISNG)** is an annual event primarily focusing on information systems and architectures. The applications of advanced information technology to such specific domains as astronomy, biology, education, geosciences, and health care are among topics of relevance to ISNG. Prototypes, designs, and tools that help the information readily flow to the user are of special interest.

ITCC (April 11-13, 2005, Las Vegas)

The **International Conference on Information Technology: Coding and Computing** (ITCC) is an international forum which brings together researchers and practitioners working on different aspects of Information Technology. It is a technical congregation where the latest theoretical and technological advances on Information Technology are presented and discussed.

The conference themes include such topics as: Information Retrieval, Operating Systems, Networks, Image/Video Processing, Digital Library, Coding & Data Compression, Watermarking, Simulation, Computer Graphics, Education, Curriculum & Accreditation, Information Assurance and Security, Data Mining and plethora of IT related technologies.

NSCEE History - an image from our past

The Convex C220

In August, 1992, the Convex Computer Corporation of Richardson, TX donated the C220 parallel/vector supercomputer, valued at $1.4 million and pictured left, to UNLV's NSCEE. The C220, with its two CPUs with a peak performance of 50 MFLOPs each (millions of floating-point operations completed per second), 256 MB of physical memory, 16 GB of high performance magnetic hard disk storage, was considered 200 to 1,000 times as powerful as the typical personal computer found on campus at the time.

The C220 benefited several ongoing computational research projects in UNLV's departments of chemistry, mechanical engineering, physics, civil engineering, electrical engineering, nursing, and mathematical sciences, as well as UNLV's Harry Reid Center for Environmental Studies, the Desert Research Institute, the State of Nevada Nuclear Projects Office, and the University of Nevada, Reno.

Articles Invited

The National Supercomputing Center for Energy and the Environment invites you to contribute articles on your work on high-performance computers and especially our resources. Please submit your articles to:

TeraWord
UNLV/NSCEE
4505 Maryland Parkway
Box 454028
Las Vegas, NV 89154-4028

email
teraword@nscee.edu

Phone
(702) 895-4153

Fax
(702) 895-4156

TeraWord is published by the National Supercomputing Center for Energy and the Environment. Materials of interest in the newsletter may be reprinted, provided acknowledgement of the source is included. Hardware and software products mentioned in this publication are trademarks of their respective companies. The use of their names does not constitute an endorsement or approval by the NSCEE or the University of Nevada Las Vegas.

NSCEE

National Supercomputing Center for Energy and the Environment

High-Performance Computing and Communications in Nevada

National Supercomputing Center for Energy and the Environment
4505 Maryland Parkway, Box 454028
Las Vegas, NV 89154-4028

Visit us at **www.nscee.edu**

UNLV
UNIVERSITY OF NEVADA LAS VEGAS