Tutorial on "**Knowledge Discovery in High-Throughput Biological Domains: Introduction to Computational Biology**"
Speakers: **Igor Jurisica, Ontario Cancer Institute / University of Toronto, Canada**
Time: **September 1st, 2005, 2:40pm-4:40pm**
Place: **CK187 (Kinesiology Building)**

The sequencing of the human genome was the first step in understanding the ways in which we are wired. Although an important milestone, this genetic blueprint provides only a "parts list", and neither the information about how the human organism is actually working, nor insight into function or interactions among the ~30 thousand constitutive parts that comprise our genome. Considering that the 30 years of molecular biology research have only annotated about 10% of this gene set, and we know even less about proteins, we increasingly rely on high-throughput data generation and analysis to provide insight into the genome, proteome and interactome.

Robotics and computational biology is rapidly changing the way we formulate and test biological hypotheses. Advances in gene expression profiling by microarray and protein profiling by mass spectrometry have suggested the potential to simultaneously view all genes expressed, all subsequent protein products, and all the interacting partners of each individual protein within a biological system. We can rapidly and accurately measure the relative activity of genes and proteins in normal and diseased tissue. The technology and datasets of such profiling-based analyses will be described along with the mathematical challenges that face the mining of the resulting datasets. We will introduce diverse computational tools and approaches used to analyze, integrate and interpret high-throughput data.

Diverse computational techniques have been applied to solve biological and medical problems over the years. Increasingly, such systems face challenges that arise from the enormous increase in information complexity and volume in these domains. In addition, the pace of evolution of our understanding of underlying principles requires continuous updates to existing databases, as well as systems that support reasoning and knowledge discovery. Performing these changes manually is becoming the bottleneck of successful application of computer science to biological and medical domains.

Introduction:

Technological advances in the high-throughput (HTP) generation of genomic and proteomic biological data have outpaced the development of computational tools to effectively analyze, visualize, and interpret this information. It is becoming clear that the high dimensionality poses a serious challenge to existing knowledge-discovery and reasoning tools. Most existing tools were developed for relational types of data, which typically have millions of instances but low dimensionality. HTP biological domains involve tens of thousands of dimensions, which cause most existing tools to either fail or provide outcomes with limited usefulness. Since these domains are characterized by the diversity of representation formalisms used, the complexity and amount of information present, uncertain or missing values, and the evolution of knowledge, it is necessary that the computational tools applied are flexible and scalable.

When designing computational systems for HTP biology, the first challenge is deciding how to store, integrate, and organize the data, so it can be accessed efficiently. However, all current HTP data-generation processes lead to incomplete and noisy databases, which pose additional challenges during the analysis and interpretation. We will discuss all aspects of knowledge management, but we explicitly focus on the analysis part, as opposed to basic storage and retrieval.

Knowledge management in HTP biology:

Knowledge management is concerned with the representation, organization, acquisition, creation, use and evolution of knowledge in its many forms. Ontology is a branch of Philosophy concerned with the study of what exists. Effectively managing biological knowledge requires efficient representation schemas, flexible and scalable retrieval algorithms, robust and accurate analysis approaches and reasoning systems. We will discuss examples of how certain representation schemes support efficient retrieval and analysis, how the annotation and system integration can be supported using shareable and reusable ontologies, and how to manage tacit human knowledge.

Knowledge discovery in HTP biology:

The advent of HTP techniques in biology, generating tens of thousands of data points within each experiment have created challenging issues in data analysis. Clustering algorithms initially enabled analysis of such data, but as more heterogeneous parameters are added (e.g., clinical outcome, genotype), more sophisticated algorithms are required to aid integration, visualization and interpretation. Multiple types of noise present in the data sets make their integration challenging. We will review the main approaches, and use examples from gene expression, protein expression, protein-protein interaction, and protein crystallization data analysis.

One fallacy in dealing with HTP biological data is ascribing too much meaning to individual data points. Many such datasets, either gene expression profiles or protein-protein interactions contain noise that can prevent reliable conclusions for specific genes or proteins. Estimates of the error rate in existing protein interaction datasets run as high as 30%.

We will discuss usefulness of integration of both data and algorithms for the analysis of HTP biological datasets. We describe integration of microarray and protein-protein interaction data, integration of phenotypic, functional and structural databases, integration of clustering and graph-theory based analysis of diverse protein-protein interaction databases.

Conclusions:

It is becoming clear that the post-genomic era is providing more computational opportunities and challenges than originally expected. Computational biology is changing from being a trendy extension for biological research, to becoming mainstream necessary component of it.

Systems biology aims at studying biological systems as a whole, not just parts, which includes understanding their structure, processes and control in a detail to enable understanding of behaviors and simulation of causal effects. This approach uses systematic and system-based approaches. Besides integrated approach, systems biology also calls for computational hypothesis generation and simulation of processes and organisms.

Understanding normal and disease states of any organism requires integrated and systematic approach. We lack understanding, and we are ramping up technologies to produce vast amounts of genomic and proteomic data. This provides both the opportunity and a challenge. No single database or algorithm will be successful at solving complex analytical problems. Thus, we need to integrate different tools and approaches, multiple single data type repositories, and repositories comprising diverse data types. We will discuss the move from computational biology to systems biology to cancer informatics.

About the speaker:

Igor Jurisica is a Scientist at the Ontario Cancer Institute, University Health Network since 2000, and Assistant Professor in the Departments of Computer Science and Medical Biophysics, University of Toronto, Adjunct Assistant Professor at School of Computing Science, Queen's University, and a Visiting Scientist at the IBM Centre for Advanced Studies.
He received a PhD degree in 1998 from the University of Toronto and MSc. degrees in Electrical Engineering from the Slovak Technical University and in Computer Science from the University of Toronto in 1991 and 1993 respectively.
His research focuses on computational biology, and representation, analysis and visualization of high dimensional data generated by high-throughput biology experiments. Of particular interest is the use of comparative analysis in the mining in integrating of different dataset types such as protein-protein interaction, gene expression profiling, and high-throughput screens for protein crystallization.

Tutorial on "**Intelligent signal processing in multimedia**"
Speakers: **Andrzej Czyzewski and Bozena Kostek, Gdansk University of Technology, Poland**
Time: **September 1st, 2005, 5pm-7pm**
Place: **CK187 (Kinesiology Building)**

Part I. Perceptual bases of multimedia technology and systems:

A huge amount of digitized audiovisual information is available in databases and archives, and in the Internet. The value of information can be evaluated in terms of data management. Processes such as automatic retrieval, access, translation, conversion, and filtration are nowadays standardized. MPEG-7 and MPEG-21 standards, understood as multimedia content description and management, are presented and discussed within the tutorial framework. The main characteristics of these standards are discussed. Issues related to low- and high-level of audiovisual content are described. In addition, the interoperability with other metadata standards already envisioned is pointed out. Also, perceptual bases of multimedia technology are presented in the context of modeling human auditory system and music cognitive bases. A few case-studies are shown and discussed on the basis of applications implemented at the Multimedia Systems Department of Gdansk University of Technology.

Part II. Applications:

Typical applications of computer technologies in multimedia only rarely consider the opportunities of data processing with the use of methods, which stem from artificial intelligence. A contrary situation can be observed in speech acoustics, where artificial intelligence methods are commonly used for automatic speech recognition and speakers identification. What is essential here, is the fact that methods of analysis and signal processing developed on the basis of speech acoustics have not been transferred respectively so far to other related areas, e.g. as a method of intelligent analysis and processing of the audio signal. In the meantime the area of multimedia has an extensive demand for applications of intelligent signal processing. An important area of applications of artificial intelligence algorithms is also analysis of results of subjective assessments; another area is related to experimental applications to automatic decision making related to processing of feature vectors sets representing audio and video.

Research experiments currently being carried out at the Multimedia Systems Department of the Gdansk University of Technology, encompass the implementation of selected artificial intelligence methods for acquisition and musical signal identification purposes, for restoration of old recordings and for applying these methods to verification of subjective evaluations in acoustics. The problems posed here are being resolved first of all with the use of the rough set method and algorithms of neural networks and also with the use of fuzzy logic. Moreover intelligent signal processing finds applications in such domains as: environment monitoring, hearing aids, homeland security systems (biometric features identification) and many others.

The tutorial will demonstrate a summarized presentation of some up-to-date applications of intelligent data analysis, which were studied at the Multimedia Systems Department of the Gdansk University of Technology. On the example of research experiments being carried out, the tutorial will present possibilities, which result from replacing traditional methods of analysis and signal and image processing by intelligent algorithms.

About the speakers:

Andrzej Czyzewski is a native of Gdansk, Poland. He received his M.Sc. degree in Sound Engineering from the Gdansk University of Technology in 1982, his Ph.D. degree in 1987 and his D.Sc. degree in 1992 from the Cracov Academy of Mining and Metallurgy. Prof. Czyzewski joined the staff of the Sound Engineering Department of the Gdansk University of Technology in 1984. In December 1999 Mr. President of Poland granted him the title of Professor. In 2002 the Senate of his University approved him to the position of Full Professor.

His main interests are multimedia technology with focus on interactive & intelligent multimedia applications; digital signal processing with focus on soft computing algorithms (neural networks, fuzzy logic and rough sets.); digital studio recording technology (sound & vision); biomedicine with focus on

applications to hearing, speech & vision; database technology with focus to intelligent data query; military computer systems with focus to air force applications.

Currently he is the Head of the Multimedia Systems Department of the Technical University of Gdansk; Director of the Doctoral Studies at the Faculty of Electronics, Telecommunications and Informatics. He is a Committee member of this Section. He is also a member of the Acoustic Committee of the Polish Academy of Sciences, and also member of: IEEE, International Rough Set Society, and International Fluency Assoc. He acts also as a member of the Scientific Council of the Institute of Physiology & Pathology of Hearing in Warsaw. Since 2002 he works for the committee for distance learning technology implementation at his University (the Gdansk University of Technology).

In 2000 Mr. Prime Minister of Poland presented him his First Prize for achievements in technical sciences. In the years 2000-2002 some of projects guided by Prof. Czyzewski were nominated to the Stockholm Challenge Award, to the Europrix'2000 prize (Europe's Best in Multimedia), were distinguished by the Polish Business Club, awarded gold and silver medals on the international invention exhibitions and represented Poland at various European IST (Information Society Technology) events also outside Europe. The Audio Engineering Society presented him its Fellowship for "outstanding achievements in soft computing applications".

He is author of more than 300 research papers published in international journals and presented in congresses & conferences around the World. In 1991, Prof. Czyzewski published a monograph devoted to digital audio operations; in 1998 he published in Poland his book entitled "Digital Sound" which won him the Prize of Ministry of Education of Poland. He is also author of 8 Polish patents in the domain of computer science and 4 international patent applications in the domain of telemedicine.

Bozena Kostek received her M.Sc. degree in Sound Engineering from the Technical University of Gdansk. Following this, she completed her studies at the Institute of Management and Organization and received a second M.Sc. from the Technical University of Gdansk. As a part of her studies, she pursued her interest in journalism and took some courses in this area. She played piano for a number of years. From 1987-1989, Dr. Kostek studied at the Paul Sabatier University in Toulouse, France, receiving the Diplome d'Acoustique.

In 1992, she supported her thesis devoted to the quality of the pipe organ control systems and received her Ph.D. degree from the Technical University of Gdansk. In March 2000 she supported her D.Sc. degree at the Institute of Research Systems of the Polish Academy of Sciences in Warsaw. Presently, Dr. Kostek is an Associate Professor with the Multimedia Systems Department. In 1996 she obtained the M. Kwiek Award for the outstanding scientific presentation from the Acoustical Society of Poland, and in 1995 and 1996 two awards from the Polish Electrotechnical Society. She was also awarded several times by the President of the Technical University of Gdansk for her scientific achievements. In November 2000 she obtained an award for scientific achievements from Mr. Prime Minister of Poland and in 2002 from the Ministry of Health for coauthoring telemedical systems.

She led a number of research projects sponsored by the Polish State Committee for Scientific Research. The subjects of the mentioned projects concern new methods of musical instrument sound recognition based on soft computing approach, pipe organ control based on fuzzy-logic, acoustic sound quality investigations, digital processing algorithms for hearing aids and cochlear implants and others. Dr Kostek has presented more than 250 scientific papers in journals and at conferences. In 1999 she published a book entitled "Soft Computing in Acoustics", and later in 2002 she she co-authored a book devoted to "Computer technology applications to audiology and speech therapy".

Her teaching responsibilities include classes in psychophysiology of hearing, studio technology, sound reinforcement, musical acoustics, electroacoustical measurements. She was also supervisor of approx. 80 Master theses, and several Ph.D. theses.

In 1991, she helped to form the Polish Section of the Audio Engineering Society, and since that time has served as a member of the Committee. She is also a member of the Polish Acoustical Society and of the Acoustical Society of America, IEEE, Rough Set Society, also she belongs to other Polish and international scientific societies.

Her main scientific interests are musical acoustics, psychophysiology of hearing, and studio technology, as well as applications of artificial intelligence and soft computing methods to the mentioned domains.

In 2003 she was elected the Vice-President of the Audio Engineering Society for Central Europe.