Academy of Mathematics and Systems Science
Chinese Academy of Sciences

2017 AMSS Workshop on Biomedical Informatics

August 29 (Tuesday), 2017, Beijing, China

Organizer: Academy of Mathematics and Systems Science, Chinese Academy of Sciences (AMSS)
Location: Meeting Room 219, South Building of AMSS,
55 Zhong-Guan-Cun East Road, Beijing 100190, China

Sponsors:

National Key Research and Development Program of China under grant 2016YFB1000902
Beijing Municipal Science &Technology Commission under grant Z161100001616002
National Natural Science Foundation of China under grant 61621003
Institute of Computing Technology, Chinese Academy of Sciences
Key Laboratory of Management, Decision and Information Systems (MADIS), Chinese Academy of Sciences

9:00 – 12:00	**The morning session, chaired by Songmao Zhang**
9:00 – 9:50	Using lexical and structural features for quality assurance of biomedical ontologies – Application to SNOMED CT Speaker: **Olivier Bodenreider** *US National Library of Medicine, National Institutes of Health, USA*
9:50 – 10:30	A comparative study of UMLS and TCMLS Speaker: **Cungen Cao** *Institute of Computing Technology, Chinese Academy of Sciences*
10:30 – 10:40	**Coffee Break**
10:40 – 11:20	Requirements engineering for clinical decision support systems in oral medicine Speaker: **Lin Liu** *School of Software, Tsinghua University*
11:20 – 12:00	Network ontology analysis Speaker: **Yong Wang** *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*
12:00 – 14:00	**Lunch**
14:00 – 17:00	**The afternoon session, chaired by Ruqian Lu**
14:00 – 14:50	Natural language processing for biomedical applications Speaker: **Pierre Zweigenbaum** *Computer Science Laboratory for Mechanics and Engineering Sciences (LIMSI), France*
14:50 – 15:30	Digitalization at the Peking University School and Hospital of Stomatology Speaker: **Zhanqiang Cao** *Information Center, Peking University School and Hospital of Stomatology*
15:30 – 15:40	**Coffee Break**
15:40 – 16:20	Graphical features of functional genes in human protein interaction network Speaker: **Jinhu Lü** *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*
16:20 – 17:00	Matching biomedical ontologies based on formal concept analysis Speaker: **Songmao Zhang** *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

Introduction to the Talks

Talk: Using lexical and structural features for quality assurance of biomedical ontologies – Application to SNOMED CT

Abstract: SNOMED CT is the largest clinical ontology in the world, used in over 30 countries to support clinical documentation and analytics. Quality assurance of large ontological systems such as SNOMED CT is an indispensable part of their life cycle, but remains challenging. We focus on the detection of missing hierarchical relations in SNOMED CT, leveraging both lexical and structural. Using description logics (DL) to represent term inclusion, a DL classifier can infer missing hierarchical relations (between a given term and the term that includes it). We also identify non-lattice subgraphs in SNOMED CT as indicative of potential errors and leverage lexical patterns to propose remediation for these errors. Preliminary results indicate that these automated methods for quality assurance can facilitate the work of human editors.

About the speaker: Olivier Bodenreider is a Senior Scientist and Chief of the Cognitive Science Branch of the Lister Hill National Center for Biomedical Communications at the U.S. National Library of Medicine. Dr. Bodenreider is a Fellow of the American College of Medical Informatics. He received an MD degree from the University of Strasbourg, France in 1990, and a PhD in Medical Informatics from the University of Nancy, France in 1993. Before joining NLM in 1996, he was assistant professor for Biostatistics and Medical Informatics at the University of Nancy, France, Medical School. His research focuses on terminology and ontology in the biomedical domain, both from a theoretical perspective (quality assurance, interoperability) and in their application to natural language processing, knowledge discovery, and information integration.

Talk: A comparative study of UMLS and TCMLS

Abstract: In this talk, I will examine differences of relationships and types between UMLS and TCMLS (A unified language system of traditional Chinese medicine). Based on this, I will try to explicate characteristics of traditional Chinese medicine in observing and describing physiological and pathological processes and in treating diseases.

About the speaker: Cungen Cao received his PhD degree in 1993 from the Institute of Mathematics, Chinese Academy of Sciences. He is currently a full-time professor of the Institute of Computing Technology of Chinese Academy of Sciences. His research interests are large-scale knowledge acquisition, knowledge management, and knowledge-intensive applications. His research results have been put into practice in the telecom industry and financial industry of China. He also has a good reputation in the knowledge processing of traditional Chinese medicine.

Talk: Requirements engineering for clinical decision support systems in oral medicine

Abstract: We are witnessing a radical change of attitude towards health information management – the adoption of data analytics to help people monitor and predict health conditions proactively. Most hospitals are aggregating data to provide analytics functions as part of their daily practices. Medical doctors start developing data repositories of their own specialty so that evidence-based clinical decisions can be made. In this talk, I will share my observations on the status quo in this area, and argue that health analytics projects need to have its requirements engineered, answering questions such as: what data it is dealing with, what data processing procedures is required, and what are the usage purposes for the data? The end products will be a set of business and user requirements (both functional, and nonfunctional) that reveals stakeholder needs of analytics results, in terms of how quickly, how often and in what format. I will present the problems and challenges encountered, report the experience and findings in our oral medicine project.

About the speaker: Lin Liu is currently Associate Professor at the School of Software, Tsinghua University. She received her Ph.D. in Computer Science from the Institute of Mathematics, Chinese Academy of Sciences. Her interests are in the areas of requirements engineering, knowledge engineering, and services computing. Her research emphasizes concepts and techniques for modeling and systematically analyzing software systems that can reason and learning to satisfy human needs, in particular, their applications in the healthcare domain. She is serving on the Editor board of Requirements Engineering Journal with Springer, and the Journal of Software in Chinese.

Talk: Network ontology analysis

Abstract: Biological function qualitatively described by Gene Ontology system is fundamental for the current genome research. Gene ontology analysis has become a popular, standard, and important tool in bioinformatics study. The current ontology analyses are mainly conducted in individual gene or a gene list, resulted by high-throughput genomic, proteomic and bioinformatics scanning technologies. However, recent advances in network biology indicate that cellular networks offer a new conceptual framework that could potentially help us to understand biological functions beyond single gene or gene list. For example, recent molecular network analysis reveals that the same list of genes with different interactions may perform different functions. Therefore, it is necessary to consider molecular interactions to correctly and specifically annotate biological networks.

In this talk, we will introduce our recent efforts on network ontology analysis. Firstly, we propose a novel Network Ontology Analysis (NOA) method to perform gene ontology enrichment analysis on biological networks. We will show that NOA outperforms the existing enrichment analysis methods by fully utilizing the information from biological networks. Specifically we will show: (i) NOA can capture the change of functions not only in dynamic transcription regulatory networks but also in rewiring protein interaction networks while the traditional methods cannot and (ii) NOA can find more relevant and specific functions than traditional methods in different types of static networks. Secondly, we will present our efforts to further integrate the ontology space and network space to reveal the central part and functional dynamics of biological network.

About the speaker: Yong Wang is Professor at the Academy of Mathematics and Systems Science, Chinese Academy of Sciences (AMSS). He is also a joint faculty member in the National Center for Mathematics and Interdisciplinary Sciences (NCMIS), Chinese Academy of Sciences. He received his Ph.D. degree in Operations Research and Control Theory from AMSS in 2005, his Master's Degree in Operations Research and Control Theory from the Dalian University of Technology in 2002 and his Bachelor's Degree in Mathematics and Physics from the Inner Mongolia University in 1999. From 2005 to 2007, he had taken the postdoctoral position at the State Information Center in China and the Department of Electronics Information and Communications, Osaka Sangyo University in Japan. From 2007.9 to 2008.9, he visited the Bioinformatics Program in Boston University as a research associate. From 2010.10 to 2011.4, he served as the Research Staff in The Computational Biology Research Center (CBRC) of National Institute of Advanced Industrial Science and Technology (AIST) in Japan. From 2012.11 to 2016.2, he is a visiting scholar in Stanford University. His current interest is in Bioinformatics and Systems Biology.

Talk: Natural language processing for biomedical applications

Abstract: From the patient record to scientific publications, natural language has an important position in electronic health (e-Health). Exploiting information and knowledge conveyed by natural language texts raises issues of Natural Language Processing (NLP). I will exemplify applications of NLP to the analysis of text found in patient records, in medical Web sources, in the scientific biomedical literature, and in doctor-patient dialogues. I will then zoom in on methods for normalizing (also known as linking) medical concepts obtained from text, which are essential for interoperability. I will show how methods based on existing dictionaries and on machine learning can be combined to obtain state-of-the-art performance on recent datasets for the task of coding diagnoses found in death certificates into the International Classification of Diseases (ICD-10).

About the speaker: Pierre Zweigenbaum is a Senior Researcher at LIMSI (Orsay, France), a laboratory of the French National Research Council (CNRS), where he leads the ILES Natural Language Processing group. He received a PhD in Computer Science from Telecom ParisTech (1985). He then was a researcher at Assistance Publique - Paris Hospitals until 2006 when he became a CNRS researcher at LIMSI in Orsay, now part of Université Paris-Saclay. He also was a part-time professor at the National Institute for Oriental Languages and Civilizations during ten years. His research focus is Natural Language Processing (NLP), with medicine as a main application domain. His main research interests are in Information Extraction, and he is the author or co-author of methods and tools to detect various types of medical entities, expand abbreviations, resolve co-references, normalize concepts, detect relations. He has also designed methods to acquire linguistic knowledge automatically from corpora and thesauri, to help extend monolingual and bilingual lexicons and terminologies, using parallel and comparable corpora. He is Vice-President of the French association for artificial intelligence (AFIA). A former chair of the AMIA NLP Special Interest Group, he is the chair of the Francophone Special Interest Group of the International Medical Informatics Association (IMIA). He is the author of over 200 peer-reviewed conference and journal publications. He was elected fellow of the American College of Medical Informatics in 2014.

Talk: Digitalization at the Peking University School and Hospital of Stomatology

Abstract: In this talk, I will present my recent experiences in dental health information systems construction, clinical data repository and decision support system development.

About the speaker: Zhanqiang Cao is currently Chief Information Officer at the Peking University School and Hospital of Stomatology. His research interests are in the area of dental information service engineering and data engineering for population dental health. Dr. Cao was trained as a professional surgeon, whose success is mainly built upon his deep understanding to the medical domain and enthusiasm of IT adoption in medical practice.

Talk: Graphical features of functional genes in human protein interaction network

Abstract: Based on up-to-date data from various databases or literature, two large-scale HPINs and six subnetworks are constructed. We illustrate that the HPINs and most of the subnetworks are sparse, small-world, scale-free, disassortative and with hierarchical modularity. Among the six subnetworks, essential, disease and HK subnetworks are more densely connected than the others. Statistical analysis on the topological structures of the HPIN reveals that the lethal, the conserved, the HK and the TE genes are with hall-mark graphical features. Receiver operating characteristic (ROC) curves indicate that the essential genes can be distinguished from the viable ones with accuracy as high as almost 70%. Closeness, semi-local and eigenvector centralities can distinguish the HK genes from the TE ones with accuracy around 82%.

About the speaker: Jinhu Lü is a Professor with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He was a Professor and ARC Future Fellow in RMIT University, Australia and a Visiting Fellow in Princeton University, USA. His research interest includes complex networks, nonlinear circuits and systems, networked knowledge, and big data. He is an ISI Highly Cited Researcher in Engineering. Prof. Lü is also the general co-chair of the 43rd Annual Conference of the IEEE Industrial Electronics Society (IECON 2017). He received the prestigious Ho Leung Ho Lee Foundation Award, and the National Natural Science Award three times in 2008, 2012, and 2016, respectively. He is also an IEEE Fellow.

Talk: Matching biomedical ontologies based on formal concept analysis

Abstract: The goal of ontology matching is to identify correspondences between entities from different yet overlapping ontologies so as to facilitate semantic integration, reuse and interoperability. As a well developed mathematical model for analyzing individuals and structuring concepts, Formal Concept Analysis (FCA) has been applied to ontology matching (OM) tasks since the beginning of OM research, whereas ontological knowledge exploited in FCA-based methods is limited. This motivates our study introduced in this talk, i.e., to empower FCA with as much as ontological knowledge as possible for identifying mappings across ontologies. Our method, called FCA-Map, incrementally generates a total of five types of formal contexts and extracts mappings from the lattices derived, based on how concepts across ontologies share lexical tokens, is-a and part-of structures, object properties, and restrictions in axioms. Evaluation on anatomy track and large biomedical ontologies track of the 2016 Ontology Alignment Evaluation Initiative (OAEI) campaign demonstrates the effectiveness of FCA-Map and its competitiveness with 2016 OAEI top-ranked OM systems. Compared with other FCA-based OM systems, our study is more comprehensive as an attempt to push the envelope of the FCA formalism in ontology matching tasks. It’s worth mentioning that as the step-by-step process of FCA-Map goes more relying on the structures of ontologies, the mapping results became unreliable. This is understandable, as ontologies for the same domain tend to differ structurally while agree more on names. Practically, manual validation from domain experts is necessary at each step so that the mismatches do not propagate further.

About the speaker: Songmao Zhang received her PhD degree from Institute of Mathematics, Chinese Academy of Sciences (CAS) in 1992. She has been a full-time professor at Academy of Mathematics and Systems Science, CAS from 2007. Besides, she was visiting scholars in research institutions and universities in the US, Australia, Germany, France and UK. Within the area of artificial intelligence, Songmao Zhang’s research interests include ontology matching, knowledge representation and reasoning in the Semantic Web, AI-based automatic animation, data mining, and natural language understanding.