

## “生物医药自然语言处理前沿”研讨会

### “生物医药自然语言处理前沿”研讨会

### Frontier in natural language Processing and BioNLP

#### 日程表

日期	时间	内容	地点	备注	
<b>2017 年 12 月 16 日 星期六</b>	8:45-9:00	与会代表入场	逸夫楼 C603		
	9:00-9:10	开幕式			
	9:10-10:00	特邀报告一：Everything that you need to know about language and natural processing in 10 graphs (Kevin B Cohen 科罗拉多大学丹佛分校)			
	10:00-10:50	特邀报告二：Tagging and Parsing for Biomedical information (Alex Chengyu Fang 香港城市大学)			
	10:50-11:00	茶歇(Coffee Break)			
	11:00-11:40	特邀报告三：How to extract associations from biomedical data (章文 武汉大学)			
	中午	11:40-14:00	午休(Lunch Break)		
	下午	14:00-14:40	特邀报告四：An overview of semantic relations between entities (Daniela Gifu 雅西“亚历山德鲁-库扎”大学 )	逸夫楼 C603	
		14:40-15:20	特邀报告五：Cascade Word Embedding to Sentence Embedding :A Class Label Enhanced Approach to Phenotype (袁晓辉 武汉理工大学)		
		15:20-15:30	闭幕式		

The frontier in BioNLP 研讨会

华中农业大学信息学院

湖北省农业生物信息重点实验室

## 特邀报告 ( Keynotes )

---



报告人：Kevin B Cohen  
科罗拉多大学丹佛分校

### 特邀报告一： Everything that you need to know about language and natural language processing in 10 graphs

**摘要：** Why do computational biologists and biomedical informaticists do research in natural language processing? Language processing turns out to have multiple applications in biomedical research—for example, analyzing high-throughput screens, building knowledge bases, and mining information from health records. But, language itself has specific characteristics that make it an interesting computational problem in its own right. This talk will present 10 essential facts about language, showing how they relate to the

**简介：** Kevin Bretonnel Cohen is the Director of the Biomedical Text Mining Group at the University of Colorado School of Medicine, and the D’Alembert Chair in Natural Language Processing for the Biomedical Domain at the Université Paris-Saclay (OR Paris-Saclay University). His book Biomedical natural language processing, written with Dina Demner-Fushman, is the standard text on the subject. Kevin’s work covers both theoretical topics, such as lexical semantics, and practical applications, such as predicting pediatric epilepsy surgery candidates.



---

报告人：Alex Chengyu Fang  
香港城市大学

---

## 特邀报告二：Tagging and Parsing for Bio-medical Information

**摘要：** This talk is a description of corpus annotation technologies including grammatical tagging and syntactic parsing. In particular, it will aim to discuss the importance of syntactically rich information and its relevance to high-quality retrieval of biomedical information. The talk will then describe some of the recent research at the Dialogue Systems Group, City University of Hong Kong, which is aimed at the development of an event extraction system based on deep parsing in general and fine-grained verb annotation in particular.

**简介：** Alex Chengyu Fang got his PhD in linguistics from University College London and held various academic positions there before joining City University of Hong Kong, where he is Associate Professor and lectures on corpus linguistics, cognitive linguistics and stylistics. He is Adjunct Professor at the Beijing University of Aeronautics and Astronautics. His research interests include syntactic parsing, term management, feature selection, event extraction, and dialogue act analysis. His major publications include *English Corpora and Automated Grammatical Analysis* (2007, The Commercial Press) and *Text Genres and Registers: The Computation of Linguistic Features* (2015, Springer). Alex is expert member of the following organisations and research labs: International Organization for Standardization, China National Technical Committee for Standardization of Terminologies and Language Resources, Hubei Provincial Research Laboratory in Bioinformatics, and Beijing Municipal Research Centre in Language Strategy and Policy.



报告人：章文  
武汉大学 计算机学院

### 特邀报告三： How to mine associations from biomedical data

**摘要：** The development of biology, medicine and highthroughput sequencing has generated a large number of biochemical data. Mining unobserved or undiscovered associations, such as drug-disease associations and lncRNA-disease associations, is one important issue in the biomedical data mining. We studied how to mine associations from biomedical data and design the computational methods. Our methods have good performances for several hotspot problems, and demonstrate the great potential of predicting novel associations.

**简介：** Wen Zhang, Ph.D., Associate professor from Computer School of Wuhan University. His research interests include recommender systems, network missing link prediction, matrix factorization, semi-supervised learning and applications of machine learning methods to biological data and problems. He published more than 40 papers on peer-reviewed journals and conferences.

### 特邀报告四： An Overview of Semantic Relations between Entities



报告人：Daniela Gifu  
“亚历山德鲁-库扎”大学

**摘要：** Researchers in the natural language processing (NLP) field have found impressive results in deploying text mining methods to (semi)-automated tasks such as defining, discovering and extracting semantic relations between different kind of entities (e.g. medical). At the application level, semantic relations can support various NLP tasks requiring a (lightweight or heavyweight) interpretation of meaning. A key challenge is content analysis to organize, analyse and extract concepts from large amounts of data using different tools or text mining algorithms (e.g. MetaMap tool focused on biomedical text). Another challenge is text classification in order to define, for instance, entities types based on features sets (e.g. medical entities such as Disease or Drug). Moreover, many studies are based on semantic relations extracted from the biomedical data using different Question Answering (QA) systems (e.g. SemRep focused in MEDLINE citations).

**简介：** Dr. Daniela GIFU is a researcher in the NLP-Group@UAIC-FII, Faculty of Computer Science, “Alexandru Ioan Cuza” University of Iasi (UAIC), Romania and the Institute for Theoretical Computer Science, Romanian Academy - Iași branch, being, since 2013, an affiliated scientific researcher at the Center for Advanced Research in Applied Informatics, University of Craiova, Romania. Her main research interests include Natural Language Processing tasks, most of them in correlation with discourse analysis. She has (co-)authored several books and journal articles, and more than 100 conference papers, many of them focused on problems of semantic and pragmatics analysis. She was a member of the Organizing Committees of the 23 International Conferences and member of the Scientific PC of the 33th International Conferences such as ACL, JCDL, IJCAI, JCDL, LREC. She has, also, a prolific literary activity: editor in chief at the Literary Destinies magazine, Montreal, Canada, editor in chief at the Union of Professional Journalists Magazine, București, Romania, member of Romanian Writers Association of Canada, vice president of the Union of Professional Journalists of Romania.



---

报告人：袁晓辉  
武汉理工大学

---

## 特邀报告五：Cascade Word Embedding to Sentence Embedding: A

### Class Label Enhanced Approach to Phenotype Extraction

**摘要：** In molecular biology, phenotypes are often described using complex semantics and diverse biomedical expressions, thereby facilitating the development of named entity recognition (NER). Here, we propose a novel approach of recognizing plant phenotypes by cascading word embedding to sentence embedding with a class label enhancement. We utilized a word embedding method to find high-frequency phenotypes with original sentences used as input in a sentence embedding method. Using this cascaded approach, we identified author-specific phenotypic expressions. In addition, we integrated a negative class label enhanced (NCLE) algorithm into our method to further optimize the training model of Sen2Vec. We used 56,748 PubMed abstracts of model organism *Arabidopsis thaliana* to test the effectiveness of our approach, which results in a 135% increase in the number of new phenotypic descriptions compared with the original phenotype ontology.

**简介：** Xiaohui Yuan is a professor in Department of Computer Science, Wuhan University of Technology, China. He received master in Computer Science from Wuhan University of Technology in 2002 and Ph.D. degree from Hokkaido University in Applied Mathematics in 2007. He did a postdoc research at Hokkaido University from 2007-2010. He joined the Key Lab of Soybean Molecular Breeding, Chinese Academy of Science from 2010, as a leader of bioinformatics team. In 2016, Dr. Yuan moved to Wuhan University of Technology. His research concerns Artificial Intelligence and Bioinformatics. Especially text mining, reasoning and deep learning methods for integrating multi-omics data. He has published over 30 papers in scientific journals including: Nature genetics, Genome Biology, Bioinformatics, Physical Review E and Chaos.