Introduction

In the era of machine learning and AI, the importance of data, either for training or for evaluating the learning-based models, is more and more evident. The performance of any AI device is greatly affected by the data it is trained on. Also, it is common that an evaluated performance of an AI device based on one benchmark data set is critically different when evaluated on another data set. This is particularly true for text mining where the performance strongly depends on various factors: the goal or the type and the language of the documents. These examples dictates the importance of sharing benchmark datasets, so that evaluation results can be comparable to each other. It is a key for efficient advancement of the technology.

BioNLP Open Shared Tasks is organized to promote the sharing of computational tasks of biomedical text mining and also solutions to them. Here sharing a task means sharing benchmark datasets and evaluation systems. It is a continuation of the previous efforts organized around the BioNLP Shared Task (BioNLP-ST) workshop series (2009, 2011, 2013, 2016).

This year, six tasks are contributed by voluntary task organizers. Two tasks, the Bacteria-Biotope (BB) and SeeDev tasks, are a continuation of their previous editions. BB targets the extraction of information about bacterial biotopes and phenotypes, while SeeDev focuses on extracting events of genetic and molecular mechanisms involved in plant seed development. The PharmaCoNER task is a named entity recognition task for pharmacological substances, compounds and proteins. Particularly it targets Spanish texts, which brings the new challenge of dealing with multilingualism. The CRAFT task is presented as a highly challenging task, aiming at annotating texts with rich semantics, and a full stack of linguistic structures. AGAC proposes to extract compositional concepts for drug repurposing. Finally RDoc is an Information Retrieval task in the field of neuroscience.

For the six tasks, a total of 45 teams participated. For the workshop, paper submissions were open exclusively to the teams that had completed at least one task as well as the task organizers. 43 reviewers in the Program Committee selected 30 papers to be presented for the workshop out of 38 submitted papers. We are happy to present the papers and we believe it to be a rare chance to compare various tasks of biomedical text mining, and also various solutions to them.

BioNLP-OST Organizers
- Jin-Dong Kim, DBCLS
- Claire Nédellec, INRA
- Robert Bossy, INRA
- Louise Deléger, INRA
Organizing Committee:

Jin-Dong Kim, DBCLS, Japan
Claire Nédellec, INRA, France
Robert Bossy, INRA, France
Louise Deléger, INRA, France

Program Committee:

Eiji Aramaki, NAIST, Japan
Mouhamadou Ba, INRA, France
Michael Bada, University of Colorado Denver, USA
William Baumgartner, University of Colorado Denver, USA
Robert Bossy, INRA, France
Tiffany Callahan, University of Colorado Denver, USA
Thierry Charnois, LIPN Université Paris 13, France
Kevin Cohen, Univ. of Colorado, USA
Nigel Collier, Univ. of Cambridge, UK
Louise Deléger, INRA, France
Dina Demner-Fushman, NLM, USA
Richard Eckart de Castilho, UKP Technische Universität Darmstadt, Germany
Arnaud Ferré, INRA, France
Juliane Fluck, Fraunhofer SCAI, Germany
Cyril Grouin, LIMSI CNRS, France
Tianyong Hao, South China Normal Univ., China
Antonio Jimeno Yepes, IBM research, Australia
Indika Kahanda, Montana State University, USA
Jin-Dong Kim, DBCLS, Japan
Sun Kim, NCBI, USA
Martin Krallinger, Barcelona Supercomputing Center, Spain
Anastasia Krithara, Demokritos, Greece
Robert Leaman, NCBI, USA
Jake Lever, Stanford University, USA
Zhiyong Lu, NCBI, USA
Claire Nédellec, INRA, France
Mariana Neves, German Federal Institute for Risk Assessment, Germany
Arzucan Ö zgür, Bogazici University, Turkey
Fabio Rinaldi, Univ of Zurich, Switzerland
Kirk Roberts, University of Texas, USA
Min Song, Yonsei Univ., Korea
Xavier Tannier, LIMICS Sorbonne Université, France
Yuka Tateishi, NBDC, Japan
Karin Verspoor, Univ. of Melbourne, Australia
Chih-Hsuan Wei, NCBI, USA
Davy Weissenbacher, UPenn, USA
Jingbo Xia, Huazhong Agricultural Univ., China
Kaiyin Zhou, Huazhong Agricultural Univ., China
# Table of Contents

*PharmaCoNER: Pharmacological Substances, Compounds and proteins Named Entity Recognition track*
Aitor Gonzalez Agirre, Montserrat Marimon, Ander Intxaurrondo, Obdulia Rabal, Marta Villegas and Martin Krallinger ................................................................. 1

*When Specialization Helps: Using Pooled Contextualized Embeddings to Detect Chemical and Biomedical Entities in Spanish*
Manuel Stoeckel, Wahed Hemati and Alexander Mehler ...................................................... 11

*VSP at PharmaCoNER 2019: Recognition of Pharmacological Substances, Compounds and Proteins with Recurrent Neural Networks in Spanish Clinical Cases*
Víctor Suárez-Paniagua ........................................................................................................ 16

*IxaMed at PharmaCoNER Challenge 2019*
Xabier Lahuerta, Iakes Goenaga, Koldo Gojenola, Aitziber Atutxa Salazar and Maite Oronoz . 21

*NLLNDE: Enhancing Neural Sequence Taggers with Attention and Noisy Channel for Robust Pharmacological Entity Detection*
Lukas Lange, Heike Adel and Jannik Strötgen .................................................................. 26

*A Deep Learning-Based System for PharmaCoNER*
Ying Xiong, Yedan Shen, Yuanhang Huang, Shuai Chen, Buzhou Tang, Xiaolong Wang, Qingcai Chen, Jun Yan and Yi Zhou ...................................................... 33

*Deep neural model with enhanced embeddings for pharmaceutical and chemical entities recognition in Spanish clinical text*
Renzo Rivera and Paloma Martínez ..................................................................................... 38

*A Neural Pipeline Approach for the PharmaCoNER Shared Task using Contextual Exhaustive Models*
Mohammad Golam Sohrab, Minh Thang Pham, Makoto Miwa and Hiroya Takamura ........... 47

*Biomedical Named Entity Recognition with Multilingual BERT*
Kai Hakala and Sampo Pyysalo ............................................................................................. 56

*An Overview of the Active Gene Annotation Corpus and the BioNLP OST 2019 AGAC Track Tasks*
Yuxing Wang, Kaiyin Zhou, Mina Gachloo and Jingbo Xia ............................................... 62

*Trigger Word Detection and Thematic Role Identification via BERT and Multitask Learning*
Dongfang Li, Ying Xiong, Baotian Hu, Hanyang Du, Buzhou Tang and Qingcai Chen .......... 72

*DeepGeneMD: A Joint Deep Learning Model for Extracting Gene Mutation-Disease Knowledge from PubMed Literature*
Feifan Liu, Xiaoyu Zheng, Bo Wang and Catarina Kiefe .................................................... 77

*Biomedical relation extraction with pre-trained language representations and minimal task-specific architecture*
Ashok Thillaisundaram and Theodosia Togia ................................................................. 84

*RACAI’s System at PharmaCoNER 2019*
Radu Ion, Vasile Florian Păis and Maria Mitrofan .......................................................... 90
Transfer Learning in Biomedical Named Entity Recognition: An Evaluation of BERT in the Pharma-CoNER task
Cong Sun and Zhihao Yang ................................................................. 100

A Multi-Task Learning Framework for Extracting Bacteria Biotope Information
Qi Zhang, Chao Liu, Ying Chi, Xuansong Xie and Xiansheng Hua ........................... 105

junyi li, Xiaobing Zhou, Yuhang Wu and Bin Wang ........................................ 110

Using Snomed to recognize and index chemical and drug mentions.
Pilar López Úbeda, Manuel Carlos Díaz Galiano, L. Alfonso Urena Lopez and Maite Martin ... 115

Bacteria Biotope at BioNLP Open Shared Tasks 2019
Robert Bossy, Louise Deléger, Estelle Chaix, Mouhamadou Ba and Claire Nédellec .......... 121

Linguistically Informed Relation Extraction and Neural Architectures for Nested Named Entity Recognition in BioNLP-OST 2019
Pankaj Gupta, Usama Yaseen and Hinrich Schütze ........................................... 132

An ensemble CNN method for biomedical entity normalization
Pan Deng, Haipeng Chen, Mengyao Huang, Xiaowen Ruan and Liang Xu ....................... 143

BOUN-ISIK Participation: An Unsupervised Approach for the Named Entity Normalization and Relation Extraction of Bacteria Biotopes
İlknur Karadeniz, Ömer Faruk Tuna and Arzuçan Özugur ........................................ 150

Bacteria Biotope Relation Extraction via Lexical Chains and Dependency Graphs
Wuti Xiong, Fei Li, Ming Cheng, Hong Yu and Donghong Ji ..................................... 158

Integration of Deep Learning and Traditional Machine Learning for Knowledge Extraction from Biomedical Literature
Jihang Mao and Wanli Liu ................................................................. 168

CRAFT Shared Tasks 2019 Overview — Integrated Structure, Semantics, and Coreference
William Baumgartner, Michael Bada, Sampo Pyysalo, Manuel R. Ciosici, Negacy Hailu, Harrison Pielke-Lombardo, Michael Regan and Lawrence Hunter ......................... 174

UZH@CRAFT-ST: a Sequence-labeling Approach to Concept Recognition
Lenz Furrer, Joseph Cornelius and Fabio Rinaldi .................................................. 185

Coreference Resolution in Full Text Articles with BERT and Syntax-based Mention Filtering
Hai-Long Trieu, Anh-Khoa Duong Nguyen, Nhung Nguyen, Makoto Miwa, Hiroya Takamura and Sophia Ananiadou ................................................................. 196

Neural Dependency Parsing of Biomedical Text: TurkuNLP entry in the CRAFT Structural Annotation Task
Thang Minh Ngo, Jenna Kanerva, Filip Ginter and Sampo Pyysalo .............................. 206

RDoC Task at BioNLP-OST 2019
Mohammad Anani, Nazmul Kazi, Matthew Kuntz and Indika Kahanda ....................... 216
BioNLP-OST 2019 RDoC Tasks: Multi-grain Neural Relevance Ranking Using Topics and Attention Based Query-Document-Sentence Interactions

Pankaj Gupta, Yatin Chaudhary and Hinrich Schütze .............................. 227
Conference Program

November 4, 2019

09:00–09:10  Opening

Session 1: PharmaCoNER

09:10–09:30  PharmaCoNER: Pharmacological Substances, Compounds and proteins Named Entity Recognition track
Aitor Gonzalez Agirre, Montserrat Marimon, Ander Intxaurondo, Obdulia Rabal, Marta Villegas and Martin Krallinger

09:30–09:40  When Specialization Helps: Using Pooled Contextualized Embeddings to Detect Chemical and Biomedical Entities in Spanish
Manuel Stoeckel, Wahed Hemati and Alexander Mehler

09:40–09:50  VSP at PharmaCoNER 2019: Recognition of Pharmacological Substances, Compounds and Proteins with Recurrent Neural Networks in Spanish Clinical Cases
Víctor Suárez-Paniagua

09:50–10:00  IxaMed at PharmacoNER Challenge 2019
Xabier Lahuerta, Iakes Goenaga, Koldo Gojenola, Aitziber Atutxa Salazar and Maite Oronoz

10:00–10:10  NLNDE: Enhancing Neural Sequence Taggers with Attention and Noisy Channel for Robust Pharmacological Entity Detection
Lukas Lange, Heike Adel and Jannik Strötgen

10:10–10:20  A Deep Learning-Based System for PharmaCoNER
Ying Xiong, Yedan Shen, Yuanhang Huang, Shuai Chen, Buzhou Tang, Xiaolong Wang, Qingcai Chen, Jun Yan and Yi Zhou

10:30–11:00  Coffee Break
November 4, 2019 (continued)

Session 2: PharmaCoNER and AGAC

11:00–11:10  Deep neural model with enhanced embeddings for pharmaceutical and chemical entities recognition in Spanish clinical text
Renzo Rivera and Paloma Martinez

11:10–11:20  A Neural Pipeline Approach for the PharmaCoNER Shared Task using Contextual Exhaustive Models
Mohammad Golam Sohrab, Minh Thang Pham, Makoto Miwa and Hiroya Takamura

11:20–11:30  Biomedical Named Entity Recognition with Multilingual BERT
Kai Hakala and Sampo Pyysalo

11:30–11:50  An Overview of the Active Gene Annotation Corpus and the BioNLP OST 2019 AGAC Track Tasks
Yuxing Wang, Kaiyin Zhou, Mina Gachloo and Jingbo Xia

11:50–12:00  Trigger Word Detection and Thematic Role Identification via BERT and Multitask Learning
Dongfang Li, Ying Xiong, Baotian Hu, Hanyang Du, Buzhou Tang and Qingcai Chen

12:00–12:10  DeepGeneMD: A Joint Deep Learning Model for Extracting Gene Mutation-Disease Knowledge from PubMed Literature
Feifan Liu, Xiaoyu Zheng, Bo Wang and Catarina Kiefe

12:10–12:20  Biomedical relation extraction with pre-trained language representations and minimal task-specific architecture
Ashok Thillaisundaram and Theodosia Togia
Poster Presentations

12:20–14:00  **RACAI’s System at PharmaCoNER 2019**
Radu Ion, Vasile Florian Păis and Maria Mitrofan

12:20–14:00  **Transfer Learning in Biomedical Named Entity Recognition: An Evaluation of BERT in the PharmaCoNER task**
Cong Sun and Zhihao Yang

12:20–14:00  **A Multi-Task Learning Framework for Extracting Bacteria Biotope Information**
Qi Zhang, Chao Liu, Ying Chi, Xuansong Xie and Xiansheng Hua

junyi li, XiaoBing Zhou, Yuhang Wu and Bin Wang

12:20–14:00  **Using Snomed to recognize and index chemical and drug mentions.**
Pilar López Úbeda, Manuel Carlos Díaz Galiano, L. Alfonso Urena Lopez and Maite Martin

12:30–14:00  **Lunch Break**

Session 3: BB and SeeDev

14:00–14:20  **Bacteria Biotope at BioNLP Open Shared Tasks 2019**
Robert Bossy, Louise Deléger, Estelle Chaix, Mouhamadou Ba and Claire Nédellec

14:20–14:40  **Linguistically Informed Relation Extraction and Neural Architectures for Nested Named Entity Recognition in BioNLP-OST 2019**
Pankaj Gupta, Usama Yaseen and Hinrich Schütze

14:40–14:50  **An ensemble CNN method for biomedical entity normalization**
Pan Deng, HaiPeng Chen, Mengyao Huang, Xiaowen Ruan and Liang Xu

14:50–15:00  **BOUN-ISIK Participation: An Unsupervised Approach for the Named Entity Normalization and Relation Extraction of Bacteria Biotopes**
İlküür Karadeniz, Ömer Faruk Tuna and Arzucan Özgür
November 4, 2019 (continued)

15:00–15:10  *Bacteria Biotope Relation Extraction via Lexical Chains and Dependency Graphs*
Wuti Xiong, Fei Li, Ming Cheng, Hong Yu and Donghong Ji

15:10–15:20  *Integration of Deep Learning and Traditional Machine Learning for Knowledge Extraction from Biomedical Literature*
Jihang Mao and Wanli Liu

15:30–16:00  *Coffee Break*

**Session 4: CRAFT and RDoC**

16:00–16:20  *CRAFT Shared Tasks 2019 Overview — Integrated Structure, Semantics, and Coreference*
William Baumgartner, Michael Bada, Sampo Pyysalo, Manuel R. Ciosici, Negacy Hailu, Harrison Pielke-Lombardo, Michael Regan and Lawrence Hunter

16:20–16:30  *UZH@CRAFT-ST: a Sequence-labeling Approach to Concept Recognition*
Lenz Furrer, Joseph Cornelius and Fabio Rinaldi

16:30–16:40  *Coreference Resolution in Full Text Articles with BERT and Syntax-based Mention Filtering*
Hai-Long Trieu, Anh-Khoa Duong Nguyen, Nhung Nguyen, Makoto Miwa, Hiroya Takamura and Sophia Ananiadou

16:40–16:50  *Neural Dependency Parsing of Biomedical Text: TurkuNLP entry in the CRAFT Structural Annotation Task*
Thang Minh Ngo, Jenna Kanerva, Filip Ginter and Sampo Pyysalo

16:50–17:10  *RDoC Task at BioNLP-OST 2019*
Mohammad Anani, Nazmul Kazi, Matthew Kuntz and Indika Kahanda

17:10–17:20  *BioNLP-OST 2019 RDoC Tasks: Multi-grain Neural Relevance Ranking Using Topics and Attention Based Query-Document-Sentence Interactions*
Pankaj Gupta, Yatin Chaudhary and Hinrich Schütze

17:20–17:40  *Discussion*