Machine Translation Summit XVII

Proceedings of The 8th Workshop on Patent and Scientific Literature Translation

http://www.aamtjapio.com/pslt2019/

20 August, 2019
Dublin, Ireland

© 2019 The authors. These articles are licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

Proceedings of
The 8th Workshop on Patent and Scientific Literature Translation

http://www.aamtjapio.com/pslt2019/

20 August, 2019
Dublin, Ireland

© 2019 The authors. These articles are licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

Preface from the co-chairs of the workshop

The Workshops on Patent and Scientific Literature Translation (PSLT), beginning as Workshops on Patent Translation (WPT), had been held biennially from 2005 to 2017 as parts of Machine Translation Summits. While patent information is still one of the major application areas of machine translation, the need for translation of different kinds of scientific literature has been increasing rapidly. The workshop covers a wide range of topics related to translation of scientific literature including patents, scientific articles, and technical reports, which have common characteristics as well as their own characteristics.

This year's workshop hosted three invited talks from various aspects: utilization of machine translation for patents/scientific literatures, transferring NLP methods across languages by machine translation techniques, and multilingual NLP in biomedical domains. Each of three invited speakers (Christof Monz, University of Amsterdam, Yohei Matsutani, Japan Patent Office, and Aurélie Névéol, LIMSI-CNRS) involves, but not limited to, multiple topics described above, and thus we believe that these invited talks give us a broader view of the state-of-the-art patent and scientific literature translation. The workshop also accepted four contributed papers that deal with interesting topics in line with current trends: three are on neural machine translation by hybrid model of data parallel approach and model parallel approach, transductive data-selection algorithms, and a multi-hop attention, respectively, and one is on how patent professionals use gist machine translation for decision making. We have organized these invited talks and scientific papers into three sessions. We hope that this workshop will contribute to mutual interaction and progress of machine translation and the fields applying machine translation.

We express our sincere appreciation to the invited speakers, the authors of the contributed papers, the Program Committee Members of this workshop, and organizing members of the MT Summit 2019.

Workshop Co-chairs
Takehito Utsuro
Katsuhito Sudoh
Takashi Tsunakawa

Organizers

Program Committee Co-Chairs

Takehito Utsuro	University of Tsukuba, Japan
Katsuhito Sudoh	Nara Institute of Science and Technology, Japan
Takashi Tsunakawa	Shizuoka University, Japan

Program Committee Members

Hailong Cao	Harbin Institute of Technology, China
Key-Sun Choi	Korea Advanced Institute of Science and Technology, Republic of Korea
Chenhui Chu	Osaka University, Japan
Hiroshi Echizen'ya	Hokkai-Gakuen University, Japan
Terumasa Ehara	Ehara NLP Research Laboratory, Japan
Isao Goto	NHK (Japan Broadcasting Corporation), Japan
Eduard Hovy	Carnegie Mellon University, USA
Kenji Imamura	National Institute of Information and Communications Technology (NICT), Japan
Satoshi Kinoshita	Japan Patent Information Organization (Japio), Japan
Sadao Kurohashi	Kyoto University, Japan
Toshiaki Nakazawa	The University of Tokyo, Japan
Takashi Ninomiya	Ehime University, Japan
Bruno Pouliquen	World Intellectual Property Organization
Svetlana Sheremetyeva	South Ural State Universtiy, Russia
Jun'ichi Tsujii	National Institute of Advanced Industrial Science and Technology (AIST), Japan
Shoichi Yokoyama	Yamagata University, Japan
Jiajun Zhang	Chinese Academy of Sciences, China

Contents

Christof Monz **1**
Invited Talk: Neural Machine Translation for Dynamic Domains

Yohei Matsutani **2**
Invited Talk: Utilization of Machine Translation in the Japan Patent Office toward improvement of accessibility for Patent Information

Aurélie Névéol **3**
Invited Talk: Biomedical Natural Language processing in multiple languages: contribution of multilingual corpus and machine translation

Junya Ono, Masao Utiyama, Eiichiro Sumita **4**
Hybrid Data-Model Parallel Training for Sequence-to-Sequence Recurrent Neural Network Machine Translation

Alberto Poncelas, Gideon Maillette de Buy Wenniger, Andy Way **13**
Transductive Data-Selection Algorithms for Fine-Tuning Neural Machine Translation

Shohei Iida, Ryuichiro Kimura, Hongyi Cui, Po-Hsuan Hung, Takehito Utsuro and Masaaki Nagata **24**
A Multi-Hop Attention for RNN based Neural Machine Translation

Mary Nurminen **32**
Decision-making, Risk, and Gist Machine Translation in the Work of Patent Professionals

Invited Talk:
Neural Machine Translation for Dynamic Domains

Christof Monz
University of Amsterdam

Profile

Christof Monz is an associate professor in computer science at the Informatics Institute, University of Amsterdam. His research interests lie in the area of multilingual natural language processing and machine translation in particular. Prior to joining the University of Amsterdam he worked as a lecturer at Queen Mary University of London and as a post-doctoral research fellow at the University of Maryland Institute for Advanced Computer Studies (UMIACS). He received a PhD in Computer Science from the University of Amsterdam in 2003.

Invited Talk:
Utilization of Machine Translation in the Japan Patent Office toward improvement of accessibility for Patent Information

Yohei Matsutani
Japan Patent Office

Profile

Yohei Matsutani has served as Deputy Director at the Patent Information Policy Planning Office, Policy Planning and Coordination Department in the Japan Patent Office since July 2018. He engages in planning policies related to the translation of patent information. He joined the JPO as a patent examiner in 2004. In his 15 years' career in the JPO, he has been involved in the patent examination of medical devices, analytical instruments.

Invited Talk:
Biomedical Natural Language processing in multiple languages: contribution of multilingual corpus and machine translation

Aurélie Névéol
LIMSI-CNRS

Profile

Aurélie Névéol is a Senior Staff Scientist at the Centre National pour la Recherche Scientifique (CNRS). She received an MSc in Linguistics in 2002 and a PhD in Computer Science in 2005. She has more than 10 years experience in biomedical Natural Language Processing Research and has addressed the analysis of biomedical text from the literature and from Electronic Health Records in French and in English. Recently, she has been focusing on clinical NLP for languages other than English. She has contributed to the development of representations of clinical information to support information extraction from EHR text, which can then be used for high throughput phenotyping. In the course of her work she has also contributed to the evaluation of research methods and workflows through her participation in the H2020 MIROR project and international evaluation campaigns such as CLEF eHealth and the biomedical task at WMT.