

1st Workshop on Natural Language Processing for Programming (NLP4Prog 2021)

Held online due to COVID-19

Bangkok, Thailand
6 August 2021

ISBN: 978-1-7138-3383-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2021) by the Association for Computational Linguistics and
the Asian Federation of Natural Language Processing
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Code to Comment Translation: A Comparative Study on Model Effectiveness & Errors</i> Junayed Mahmud, Fahim Faisal, Raihan Islam Arnob, Antonios Anastasopoulos and Kevin Moran	1
<i>ConTest: A Unit Test Completion Benchmark featuring Context</i> Johannes Villmow, Jonas Depoix and Adrian Ulges	17
<i>CommitBERT: Commit Message Generation Using Pre-Trained Programming Language Model</i> Tae Hwan Jung	26
<i>Time-Efficient Code Completion Model for the R Programming Language</i> Artem Popov, Dmitrii Orekhov, Denis Litvinov, Nikolay Korolev and Gleb Morgachev	34
<i>CoTexT: Multi-task Learning with Code-Text Transformer</i> Long Phan, Hieu Tran, Daniel Le, Hieu Nguyen, James Annibal, Alec Peltekian and Yanfang Ye	40
<i>DIRECT : A Transformer-based Model for Decompiled Identifier Renaming</i> Vikram Nitin, Anthony Saieva, Baishakhi Ray and Gail Kaiser	48
<i>Shellcode_IA32: A Dataset for Automatic Shellcode Generation</i> Pietro Liguori, Erfan Al-Hossami, Domenico Cotroneo, Roberto Natella, Bojan Cukic and Samira Shaikh	58
<i>Reading StackOverflow Encourages Cheating: Adding Question Text Improves Extractive Code Generation</i> Gabriel Orlanski and Alex Gittens	65
<i>Text-to-SQL in the Wild: A Naturally-Occurring Dataset Based on Stack Exchange Data</i> Moshe Hazoom, Vibhor Malik and Ben Bogin	77
<i>Bag-of-Words Baselines for Semantic Code Search</i> Xinyu Zhang, Ji Xin, Andrew Yates and Jimmy Lin	88