HAL
open science

Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)

Thiago Castro Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van Der Lee, Simon Mille, Diego Moussallem, Anastasia Shimorina

To cite this version:

Thiago Castro Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van Der Lee, Simon Mille, et al.. Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+). WebNLG+: 3rd Workshop on Natural Language Generation from the Semantic Web, Dec 2020, Virtuel, Ireland. hal-03562633

HAL Id: hal-03562633

https://hal.science/hal-03562633v1

Submitted on 22 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

WebNLG 2020

Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)

18 December 2020
Dublin, Ireland (Virtual)

Sponsors

The WebNLG+ challenge was supported by the XNLG AI Chair on Multi-lingual, Multi-Source Text Generation, an initiative funded by the French National Research Agency, Facebook AI Research and the Great East Region (Gardent; award ANR-20-CHIA-0003, XNLG "Multi-lingual, Multi-Source Text Generation").

©2020 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-952148-59-0

Preface

Welcome to the third edition of the "Workshop on Natural Language Generation from the Semantic Web" (WebNLG+).

Initiated in 2015 and tightly linked to the eponymous WebNLG Challenge, the WebNLG workshops target two goals: to investigate the micro-planning tasks involved in generating from RDF triples (document structuring, referring expression generation, aggregation, lexicalisation, surface realisation); and to explore the potential of NLG to verbalise the Semantic Web.

Different from the WebNLG 2017 challenge, WebNLG+ includes not one but four tasks: Generating from RDF triples into English and into Russian; and parsing English and Russian text into a set of RDF triples. The challenge attracted participation from 16 teams, two of which subsequently withdrew yielding a final total of 15 submitted systems and 14 participants: 6 from the private sector (Amazon Shanghai AI Lab with two systems, Facebook, Google, Huawei Noah's Ark Lab, Orange, Sber AI Lab) and 8 from universities or public research labs (Montréal, Prague, Catalunya, São Paulo, Ireland, Ohio State, AIST). The proposed approaches include template-based methods, mono-task neural approaches, bilingual approaches and multi-task bi-directional, bilingual models. Pre-training is widely used and various data transformation (delexicalisation) and data augmentation techniques were explored.

The technical program of the WebNLG+ 2020 workshop consists of 2 long papers, 4 short papers and 15 WebNLG+ system descriptions, all of which will be presented COVID-style, either through live videos or through pre-recorded videos. The program also includes an overview of the WebNLG+ 2020 participants and the results, which will be presented by the challenge organisers.

WebNLG+ (workshop and challenge) was organised by an international group of young and less young researchers: Thiago Castro Ferreira (Federal University of Minas Gerais, Brazil), Claire Gardent (CNRS / LORIA, Nancy, France), Nikolai Ilinykh (University of Gothenburg, Sweden), Chris van der Lee (Tilburg University, The Netherlands), Simon Mille (Universitat Pompeu Fabra, Barcelona, Spain), Diego Moussallem (Paderborn University, Germany) and Anastasia Shimorina (Université de Lorraine/LORIA, Nancy, France).

We would like to thank all the people who contributed to this workshop: the authors for their submissions; the program committee for their prompt and effective reviewing; the INLG 2020 organising committee, especially the workshops chairs, Daniel Braun; the publication chair, Laura Perez-Beltrachini, and the Programme chairs, Brian Davis, Yvette Graham, John Kelleher and Gowri Sripada Pighin; and the workshop participants for sharing time and thoughts on this increasingly important research topic. Special thanks go to Yannick Parmentier (Université de Lorraine, Nancy, France) for the short notice processing of these proceedings into an ACL anthology compliant format.

Wishing you all an enjoyable, informative read!

Thiago, Claire, Nikolai, Chris, Simon, Diego and Anastasia

Committees

Program Committee

- Jose Maria Alonso, University of Santiago de Compostela, Spain
- Emmanuel Ayodele, Peoples' Friendship University of Russia, Russia
- Anja Belz, University of Brighton, UK
- Bernd Bohnet, Google Research, The Netherlands
- Thiago Castro Ferreira, Federal University of Minas Gerais, Brazil
- Benoît Crabbé, University of Paris, France
- David Dale, Independent researcher, Russia
- Stamatia Dasiopoulou, Independent researcher, Spain
- Henry Elder, ADAPT, Dublin City University, Ireland
- Claire Gardent, CNRS/LORIA, Nancy, France
- Sebastian Gehrmann, Google Research, USA
- David M. Howcroft, Heriot-Watt University, UK
- Nikolai Ilinykh, University of Gothenburg, Sweden
- Guy Lapalme, RALI-DIRO, Université de Montréal, Canada
- Simon Mille, Universitat Pompeu Fabra, Barcelona, Spain
- Chris Kedzie, Columbia University, USA
- Chris van der Lee, Tilburg University, The Netherlands
- Diego Moussalem, Paderborn University, Germany
- Dang Tuan Nguyen, Saigon University, Vietnam
- Alina Petrova, University of Oxford, UK
- Abhishek V. Potnis, IIT Bombay, India
- Marco Roberti, University of Turin, Italy
- Aleksander Shvets, Pompeu Fabra University, Spain
- Anastasia Shimorina, Université de Lorraine/LORIA, Nancy, France
- Marco Antonio Sobrevilla Cabezudo, University of São Paulo, Brazil
- Ashish Upadhyay, Robert Gordon University, UK
- Leo Wanner, Pompeu Fabra University, Spain
- Xiang Yu, University of Stuttgart, Germany
- Rui Zhang, Yale University, USA

Organising committee

- Thiago Castro Ferreira, Federal University of Minas Gerais, Brazil
- Claire Gardent, CNRS/LORIA, Nancy, France
- Nikolai Ilinykh, University of Gothenburg, Sweden
- Chris van der Lee, Tilburg University, The Netherlands
- Simon Mille, Universitat Pompeu Fabra, Barcelona, Spain
- Diego Moussalem, Paderborn University, Germany
- Anastasia Shimorina, Université de Lorraine/LORIA, Nancy, France

Table of contents

I Long papers 1

A Case Study of NLG from Multimedia Data Sources: Generating Architectural Landmark Descriptions 2
Simon Mille, Petros Alvanitopoulos, Roberto Carlini Salguero, Jens Grivolla, Montserrat Marimon Felipe, Georgios Meditskos, Maria Rousi, Klearchos Stavrothanasopoulos, Spyridon Symeonidis, Stefanos Vrochidis, Leo Wanner

OWLSIZ: An isiZulu CNL for structured knowledge validation 15
Zola Mahlaza, C. Maria Keet

II Short papers 26

A General Benchmarking Framework for Text Generation 27
Diego Moussallem, Paramjot Kaur, Thiago Ferreira, Chris van der Lee, Anastasia Shimorina, Felix Conrads, Michael Röder, René Speck, Claire Gardent, Simon Mille, Nikolai Ilinykh, Axel-Cyrille Ngonga Ngomo

Controllable Neural Natural Language Generation: comparison of state-of-the-art control strategies 34
Yuanmin Leng, François Portet, Cyril Labbé, Raheel Qader

Enhancing Sequence-to-Sequence Modelling for RDF triples to Natural Text 40
Oriol Domingo, David Bergés, Roser Cantenys, Roger Creus, José A. R. Fonollosa

Utilising Knowledge Graph Embeddings for Data-to-Text Generation 48
Nivranshu Pasricha, Mihael Arcan, Paul Buitelaar

III System descriptions 54

The 2020 Bilingual, Bi-Directional WebNLG+ Shared Task: Overview and Evaluation Results (WebNLG+ 2020) 55
Thiago Castro Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van der Lee, Simon Mille, Diego Moussallem, Anastasia Shimorina

CycleGT: Unsupervised Graph-to-Text and Text-to-Graph Generation via Cycle Training 77
Qipeng Guo, Zhijing Jin, Xipeng Qiu, Weinan Zhang, David Wipf, Zheng Zhang

Denoising Pre-Training and Data Augmentation Strategies for Enhanced RDF Verbalization with Transformers 89
Sebastien Montella, Betty Fabre, Tanguy Urvoy, Johannes Heinecke, Lina Rojas-Barahona

$\mathcal{P}^{\in}$: A Plan-and-Pretrain Approach for Knowledge Graph-to-Text Generation 100
Qipeng Guo, Zhijing Jin, Ning Dai, Xipeng Qiu, Xiangyang Xue, David Wipf, Zheng Zhang

Improving Text-to-Text Pre-trained Models for the Graph-to-Text Task 107
Zixiaofan Yang, Arash Einolghozati, Hakan Inan, Keith Diedrick, Angela Fan, Pinar Donmez, Sonal

Gupta

Leveraging Large Pretrained Models for WebNLG 2020 **117**
Xintong Li, Aleksandre Maskharashvili, Symon Jory Stevens-Guille, Michael White

Machine Translation Aided Bilingual Data-to-Text Generation and Semantic Parsing **125**
Oshin Agarwal, Mihir Kale, Heming Ge, Siamak Shakeri, Rami Al-Rfou

NILC at WebNLG+: Pretrained Sequence-to-Sequence Models on RDF-to-Text Generation **131**
Marco Antonio Sobrevilla Cabezudo, Thiago A. S. Pardo

NUIG-DSI at the WebNLG+ challenge: Leveraging Transfer Learning for RDF-to-text generation **137**
Nivranshu Pasricha, Mihael Arcan, Paul Buitelaar

RDFjsRealB: a Symbolic Approach for Generating Text from RDF Triples **144**
Guy Lapalme

Semantic Triples Verbalization with Generative Pre-Training Model **154**
Pavel Blinov

Text-to-Text Pre-Training Model with Plan Selection for RDF-to-Text Generation **159**
Natthawut Kertkeidkachorn, Hiroya Takamura

The UPC RDF-to-Text System at WebNLG Challenge 2020 **167**
David Bergés, Roser Cantenys, Roger Creus, Oriol Domingo, José A. R. Fonollosa

Train Hard, Finetune Easy: Multilingual Denoising for RDF-to-Text Generation **171**
Zdenek Kasner, Ondrej Dusek

WebNLG 2020 Challenge: Semantic Template Mining for Generating References from RDF **177**
Trung Tran, Dang Tuan Nguyen

WebNLG Challenge 2020: Language Agnostic Delexicalisation for Multilingual RDF-to-text generation **186**
Giulio Zhou, Gerasimos Lampouras

Part I

Long papers