COLING 2022
Volume 29 (2022), No. 6

Proceedings of The Workshop on Automatic Summarization for Creative Writing

The 29th International Conference on Computational Linguistics

October 17, 2022
Gyeongju, Republic of Korea

Copyright of each paper stays with the respective authors (or their employers).

ISSN 2951-2093

Preface

In recent years, automatic text summarization has seen dramatic advances due to the development of large neural language models such as BERT, BART, and PALM. However, the majority of work in this field focuses on the domain of single-document news summarization, given the availability of datasets such as CNN/DailyMail, XSum, NewsRoom, and NYTimes, among others. While this domain is important, it suffers from several limitations in its short input lengths, its focus on literal language, and its constrained discourse structure. While efforts in multi-document summarization of news and dialog bring in additional complexity, they do not address the larger problem of building datasets for truly challenging datasets. We envision that in the near future, summarization systems will need to be equipped with the ability to:

- Process long input sequences spanning up to hundreds of pages of text
- Analyze complex discourse structure such as narrative and multi-party dialog
- Interpret figurative language and convey the salient points in the input

An equally important yet underexplored domain for text summarization is creative writing, which includes documents such as books, stories, as well as scripts from plays, TV shows, and movies. Documents in this domain are uniquely characterized by their substantial input lengths, non-trivial temporal dependencies (e.g., parallel plot threads), complex structures which often combine narrative and multi-party dialogues, and a wide variety of styles. Successfully summarizing such texts requires making literary interpretations, conveying implicit information, and heavily paraphrasing the inputs. The challenges of creative writing summarization, then, require the development of systems that utilize techniques not yet explored in the field.

This workshop aims to bring together researchers and promote exciting work in the domain of creative writing summarization, with the hope of contributing to the next generation of summarization systems. The workshop includes papers on topics required for summarizing creative text as well as papers reporting on a shared task on creative text, encompassing four sub-tasks: summarization of chapters from novels, summarization of movie scripts, summarization of prime time television transcripts, and summarization of daytime soap opera transcripts.

In addition to the published papers, the workshop features eight invited talks from researchers working in summarization: Mirella Lapata (University of Edinburgh), Asli Celikyilmaz (Meta AI), Shashi Narayan (Google AI), Greg Durrett (University of Texas, Austin), Mohit Bansal (University of North Carolina, Chapel Hill), Miguel Ballesteros (Amazon AI Labs), Lu Wang (University of Michigan), and Xiaojun Wan (Peking University).

We invite you to enjoy the workshop talks and the proceedings!

The Organizing Committee

Organizing Committee

Kathleen McKeown (Columbia University), chair
Divyansh Agarwal (Salesforce Research)
Alexander Fabbri (Salesforce Research)
Simeng Han (Yale University)
Wojciech Kryściński (Salesforce Research)
Faisal Ladhak (Columbia University)
Bryan Li (University of Pennsylvania)
Ramesh Nallapati (Amazon AI Labs)
Dragomir Radev (Yale University)
Sam Wiseman (Duke University)
Tianyi Zhang (Stanford University)

Table of Contents

IDN-Sum: A New Dataset for Interactive Digital Narrative Extractive Text Summarisation
Ashwathy T. Revi, Stuart E. Middleton and David E. Millard 1

Summarization of Long Input Texts Using Multi-Layer Neural Network
Niladri Chatterjee, Aadyant Khatri and Raksha Agarwal 13

COLING 2022 Shared Task: LED Finteuning and Recursive Summary Generation for Automatic Summarization of Chapters from Novels
Prerna Kashyap 19

TEAM UFAL @ CreativeSumm 2022: BART and SamSum based few-shot approach for creative Summarization
Rishu Kumar and Rudolf Rosa 24

Long Input Dialogue Summarization with Sketch Supervision for Summarization of Primetime Television Transcripts
Nataliia Kees, Thien Nguyen, Tobias Eder and Georg Groh 29

AMRTVSumm: AMR-augmented Hierarchical Network for TV Transcript Summarization
Yilun Hua, Zhaoyuan Deng and Zhijie Xu 36

Automatic Summarization for Creative Writing: BART based Pipeline Method for Generating Summary of Movie Scripts
Aditya Upadhyay, Nidhir Bhavsar, Aakash Bhatnagar, Muskaan Singh and Petr Motlicek 44

The CreativeSumm 2022 Shared Task: A Two-Stage Summarization Model using Scene Attributes
Eunchong Kim, Taewoo Yoo, Gunhee Cho, Suyoung Bae and Yun-Gyung Cheong 51

Two-Stage Movie Script Summarization: An Efficient Method For Low-Resource Long Document Summarization
Dongqi Pu, Xudong Hong, Pin-Jie Lin, Ernie Chang and Vera Demberg 57

CREATIVESUMM: Shared Task on Automatic Summarization for Creative Writing
Divyansh Agarwal, Alexander R. Fabbri, Simeng Han, Wojciech Kryscinski, Faisal Ladhak, Bryan Li, Kathleen McKeown, Dragomir Radev, Tianyi Zhang and Sam Wiseman 67

Conference Program

October 17th, 2022

09:00–10:30 Session 1

09:00–09:30 *Modeling and Evaluating Faithful Generation across Modalities*
Mohit Bansal

09:30–10:00 *Controllable Content Creation With Planning*
Shashi Narayan

10:00–10:20 *IDN-Sum: A New Dataset for Interactive Digital Narrative Extractive Text Summarisation*
Ashwathy T. Revi, Stuart E. Middleton and David E. Millard

10:50–12:30 Session 2

10:50–11:20 *Where did I read that? It was in a book. Challenges in Summarization of Book Chapters and Dialogues*
Miguel Ballesteros

11:20–11:50 *Long Document Summarization using Efficient Attentions and Document Structure*
Lu Wang

11:50–12:10 *Summarization of Long Input Texts Using Multi-Layer Neural Network*
Niladri Chatterjee, Aadyant Khatri and Raksha Agarwal

12:10–12:30 *COLING 2022 Shared Task: LED Finteuning and Recursive Summary Generation for Automatic Summarization of Chapters from Novels*
Prerna Kashyap

October 17th, 2022 (continued)

13:50–15:40 Session 3

13:50–14:20 *Hierarchical 3D Adapters for Long Video-to-text Summarization*
Mirella Lapata

14:20–14:40 *TEAM UFAL @ CreativeSumm 2022: BART and SamSum based few-shot approach for creative Summarization*
Rishu Kumar and Rudolf Rosa

14:40–15:00 *Long Input Dialogue Summarization with Sketch Supervision for Summarization of Primetime Television Transcripts*
Nataliia Kees, Thien Nguyen, Tobias Eder and Georg Groh

15:00–15:20 *AMRTVSumm: AMR-augmented Hierarchical Network for TV Transcript Summarization*
Yilun Hua, Zhaoyuan Deng and Zhijie Xu

15:20–15:40 *Automatic Summarization for Creative Writing: BART based Pipeline Method for Generating Summary of Movie Scripts*
Aditya Upadhyay, Nidhir Bhavsar, Aakash Bhatnagar, Muskaan Singh and Petr Motlicek

16:00–18:10 Session 4

16:00–16:30 *Towards Figurative Language Generation*
Xiajun Wan

16:30–17:00 *Discourse Aware Text Summarization*
Asli Celikyilmaz

17:00–17:20 *The CreativeSumm 2022 Shared Task: A Two-Stage Summarization Model using Scene Attributes*
Eunchong Kim, Taewoo Yoo, Gunhee Cho, Suyoung Bae and Yun-Gyung Cheong

17:20–17:40 *Two-Stage Movie Script Summarization: An Efficient Method For Low-Resource Long Document Summarization*
Dongqi Pu, Xudong Hong, Pin-Jie Lin, Ernie Chang and Vera Demberg

17:40–18:10 *Summarizing Narratives with GPT-3: Measuring the next 5 years of progress*
Greg Durrett