

# Seminar Neural Text Summarization

Julius Steen  
steen@cl.uni-heidelberg.de

October 2019

## 1 Overview

This document is a preliminary schedule for the seminar and subject to change based on the number of students and their interests.

To complete the course you need to participate regularly (no more than one unexcused absence) and be active in the discussions. Each student will also have to give a graded presentation. In addition, you need to do one of the following to complete the course:

1. Write a term paper
2. Do an implementation project
3. Give a second presentation (subject to availability)

Students are expected to read the presented papers (at most two) and hand in questions or comments about them via mail to me each Monday before 3pm. These questions will be a part of your final grade.

## 2 Getting a presentation slot

To get a presentation slot, write an email to steen@cl.uni-heidelberg.de by the 26th October with a ranked list of three presentations that you would be interested in. If you want to do two presentations in lieu of a term paper, please indicate that in your mail and list an additional two or three presentations.

The dates given for the presentations are tentative and might change based on your interests and the number of PS vs. HS participants. If there is any date where it would not be possible to give your presentation, you can indicate *at most one* session where you can absolutely not give your presentation.

## 3 Differences between HS and PS

The course can be taken both as a main seminar (HS) and as a bachelor seminar (PS). The differences for the purpose of the presentation are:

- HS presentations cover a full lecture (sixty minutes, plus thirty minutes of questions and discussion), PS presentations are half a lecture (30 minutes, plus 15 minutes of questions and discussion)
- HS presentations are generally expected to cover (at least) two papers, PS presentations only need to cover one paper
- PS papers are generally easier than HS papers.

I have tentatively marked which sessions are appropriate for either HS or PS. This is subject to change based on attendance and interests of the participants. The categorization is not binding. If you find one of the HS-papers interesting, but want to do a PS, you can tell me and I will try to reschedule accordingly. If you are doing a HS and are interested in the topics for a PS-session, you can similarly tell me. You will most likely have to add in another paper or some extensive discussion, so the workload and difficulty are still fair.

## 4 Schedule

### 23.10 - Topic Overview

*No student presentation*

### 30.10 - Introduction

*No student presentation*

**Prepare:** [Lin \[2004\]](#)

### 06.11 - Canceled

*No Seminar*

### 13.11 - Extractive Single- and Multi-Document-Summarization (PS)

**Presentation 1:** [Nallapati et al. \[2017\]](#)

**Presentation 2:** [Yasunaga et al. \[2017\]](#)

**Alternative:** [Cheng and Lapata \[2016\]](#)

### 20.11 - Abstractive Summarization with Pointer Mechanisms (PS/HS)

**Presentation 1:** [Nallapati et al. \[2016\]](#)

**Presentation 2:** [See et al. \[2017\]](#)

**Background:** [Gulcehre et al. \[2016\]](#)

For HS:

**Main Papers:** Both papers from above

**Additionally:** Compare with [Gu et al. \[2016\]](#)

## 27.11 - Datasets and Task-Variations (PS)

**Presentation 1:** [Grusky et al. \[2018\]](#)

**Presentation 2:** [Narayan et al. \[2018a\]](#)

## 04.12 - Extractive Summarization with reinforcement learning (HS)

**Main Paper:** [Zhang et al. \[2018\]](#) *or* [Narayan et al. \[2018b\]](#)

**Additionally:** An introduction to the REINFORCE-algorithm [[Williams, 1992](#)]

## 11.12 - More Architectures for Abstractive Summarization (HS)

**Main Paper:** [Paulus et al. \[2018\]](#) *or* [Celikyilmaz et al. \[2018\]](#)

**Additionally:** [Rennie et al. \[2016\]](#) with a focus on the self-critical gradient algorithm

## 18.12 - Controllability (PS/HS)

For PS:

**Presentation 1:** [Fan et al. \[2018\]](#)

**Presentation 2:** [Liu et al. \[2018\]](#)

For HS:

**Main Paper:** [Makino et al. \[2019\]](#)

**Additionally:** [Shen et al. \[2016\]](#)

## 08.01 - Pretraining for Summarization (HS)

**Paper 1:** [Zhang et al. \[2019b\]](#)

**Paper 2:** [Zhang et al. \[2019a\]](#)

**Alternatively, depending on prior knowledge of participants:** An overview over transformer architecture [[Vaswani et al., 2017](#)] and/or BERT [[Devlin et al., 2018](#)]

## **15.01 - Improving Summary Coherence (PS/HS)**

For PS:

**Presentation 1:** [Wu and Hu \[2018\]](#)

**Presentation 2:** [Gabriel et al. \[2019\]](#)

For HS:

**Main Paper** [Sharma et al. \[2019\]](#)

**Additionally:** [Wu and Hu \[2018\]](#)

## **22.01 - Factual Correctness, Background knowledge (PS)**

**Presentation 1:** [Cao et al. \[2017\]](#)

**Presentation 2:** [Amplayo et al. \[2018\]](#)

## **29.01 - Abstractive Multi-Document-Summarization (PS/HS)**

**Presentation 1:** [Lebanoff et al. \[2018\]](#)

**Presentation 2:** [Fabbri et al. \[2019\]](#)

For HS:

**Paper 1:** [Liu\\* et al. \[2018\]](#)

**Paper 2:** [Fan et al. \[2019\]](#)

## **05.02 - Discussion (no presentation)**

**Prepare:** [Kryściński et al. \[2019\]](#)

## **Backup (possibly instead of correctness/KB and/or Datasets) - Unsupervised Abstractive Summarization (HS)**

Variant 1:

**Paper 1:** [Chu and Liu \[2018\]](#)

**Paper 2:** [Isonuma et al. \[2019\]](#)

Variant 2:

**Main Paper:** [Jernite \[2019\]](#)

**Additionally as background:** [Lample et al. \[2018\]](#)

## References

- Reinald Kim Amplayo, Seonjae Lim, and Seung-won Hwang. Entity common-sense representation for neural abstractive summarization. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 697–707. Association for Computational Linguistics, 2018. doi: 10.18653/v1/N18-1064. URL <http://aclweb.org/anthology/N18-1064>.
- Ziqiang Cao, Furu Wei, Wenjie Li, and Sujian Li. Faithful to the original: Fact aware neural abstractive summarization. *CoRR*, abs/1711.04434, 2017. URL <http://arxiv.org/abs/1711.04434>.
- Asli Celikyilmaz, Antoine Bosselut, Xiaodong He, and Yejin Choi. Deep communicating agents for abstractive summarization. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1662–1675, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1150. URL <https://www.aclweb.org/anthology/N18-1150>.
- Jianpeng Cheng and Mirella Lapata. Neural summarization by extracting sentences and words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 484–494, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1046. URL <https://www.aclweb.org/anthology/P16-1046>.
- Eric Chu and Peter J. Liu. Unsupervised neural multi-document abstractive summarization. *CoRR*, abs/1810.05739, 2018. URL <http://arxiv.org/abs/1810.05739>.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018. URL <http://arxiv.org/abs/1810.04805>.
- Alexander Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir Radev. Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1074–1084, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1102. URL <https://www.aclweb.org/anthology/P19-1102>.
- Angela Fan, David Grangier, and Michael Auli. Controllable abstractive summarization. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 45–54, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-2706. URL <https://www.aclweb.org/anthology/W18-2706>.

- Angela Fan, Claire Gardent, Chloe Braud, and Antoine Bordes. Using local knowledge graph construction to scale seq2seq models to multi-document inputs, 2019. URL <https://arxiv.org/pdf/1910.08435.pdf>.
- Saadia Gabriel, Antoine Bosselut, Ari Holtzman, Kyle Lo, Asli Çelikyilmaz, and Yejin Choi. Cooperative generator-discriminator networks for abstractive summarization with narrative flow. *CoRR*, abs/1907.01272, 2019. URL <http://arxiv.org/abs/1907.01272>.
- Max Grusky, Mor Naaman, and Yoav Artzi. Newsroom: A dataset of 1.3 million summaries with diverse extractive strategies. *CoRR*, abs/1804.11283, 2018. URL <http://arxiv.org/abs/1804.11283>.
- Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. Li. Incorporating copying mechanism in sequence-to-sequence learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1631–1640, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1154. URL <https://www.aclweb.org/anthology/P16-1154>.
- Caglar Gulcehre, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio. Pointing the unknown words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 140–149, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1014. URL <https://www.aclweb.org/anthology/P16-1014>.
- Masaru Isonuma, Junichiro Mori, and Ichiro Sakata. Unsupervised neural single-document summarization of reviews via learning latent discourse structure and its ranking. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2142–2152, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1206. URL <https://www.aclweb.org/anthology/P19-1206>.
- Yacine Jernite. Unsupervised text summarization via mixed model back-translation. *arXiv preprint arXiv:1908.08566*, 2019.
- Wojciech Kryściński, Nitish Shirish Keskar, Bryan McCann, Caiming Xiong, and Richard Socher. Neural Text Summarization: A Critical Evaluation. *arXiv e-prints*, art. arXiv:1908.08960, Aug 2019.
- Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc’Aurelio Ranzato. Unsupervised machine translation using monolingual corpora only. In *International Conference on Learning Representations*, 2018. URL <https://openreview.net/forum?id=rkYTTf-AZ>.
- Logan Lebanoff, Kaiqiang Song, and Fei Liu. Adapting the neural encoder-decoder framework from single to multi-document summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Pro-*

- cessing, pages 4131–4141. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/D18-1446>.
- Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/W04-1013>.
- Peter J. Liu\*, Mohammad Saleh\*, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, and Noam Shazeer. Generating wikipedia by summarizing long sequences. In *International Conference on Learning Representations*, 2018. URL <https://openreview.net/forum?id=HygOvbWC->.
- Yizhu Liu, Zhiyi Luo, and Kenny Zhu. Controlling length in abstractive summarization using a convolutional neural network. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4110–4119. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/D18-1444>.
- Takuya Makino, Tomoya Iwakura, Hiroya Takamura, and Manabu Okumura. Global optimization under length constraint for neural text summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1039–1048, Florence, Italy, July 2019. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/P19-1099>.
- Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Çağlar Gulçehre, and Bing Xiang. Abstractive text summarization using sequence-to-sequence RNNs and beyond. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 280–290, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/K16-1028. URL <https://www.aclweb.org/anthology/K16-1028>.
- Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents, 2017. URL <https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14636>.
- Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don’t give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807. Association for Computational Linguistics, 2018a. URL <http://aclweb.org/anthology/D18-1206>.
- Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Ranking sentences for extractive summarization with reinforcement learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1747–1759. Association for Computational Linguistics, 2018b. URL <http://aclweb.org/anthology/N18-1158>.

- Romain Paulus, Caiming Xiong, and Richard Socher. A deep reinforced model for abstractive summarization. In *International Conference on Learning Representations*, 2018. URL <https://openreview.net/forum?id=HkAC1QgA->.
- Steven J. Rennie, Etienne Marcheret, Youssef Mroueh, Jerret Ross, and Vaibhava Goel. Self-critical sequence training for image captioning. *CoRR*, abs/1612.00563, 2016. URL <http://arxiv.org/abs/1612.00563>.
- Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083. Association for Computational Linguistics, 2017. doi: 10.18653/v1/P17-1099. URL <http://www.aclweb.org/anthology/P17-1099>.
- Eva Sharma, Luyang Huang, Zhe Hu, and Lu Wang. An Entity-Driven Framework for Abstractive Summarization. *arXiv e-prints*, art. arXiv:1909.02059, Sep 2019. URL <https://arxiv.org/abs/1909.02059>.
- Shiqi Shen, Yong Cheng, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. Minimum risk training for neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1683–1692, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1159. URL <https://www.aclweb.org/anthology/P16-1159>.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc., 2017. URL <http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>.
- Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3):229–256, May 1992. ISSN 1573-0565. doi: 10.1007/BF00992696. URL <https://doi.org/10.1007/BF00992696>.
- Yuxiang Wu and Baotian Hu. Learning to extract coherent summary via deep reinforcement learning. *CoRR*, abs/1804.07036, 2018. URL <http://arxiv.org/abs/1804.07036>.
- Michihiro Yasunaga, Rui Zhang, Kshitijh Meelu, Ayush Pareek, Krishnan Srinivasan, and Dragomir R. Radev. Graph-based neural multi-document summarization. *CoRR*, abs/1706.06681, 2017. URL <http://arxiv.org/abs/1706.06681>.
- Haoyu Zhang, Yeyun Gong, Yu Yan, Nan Duan, Jianjun Xu, Ji Wang, Ming Gong, and Ming Zhou. Pretraining-based natural language generation for



text summarization. *CoRR*, abs/1902.09243, 2019a. URL <http://arxiv.org/abs/1902.09243>.

Xingxing Zhang, Mirella Lapata, Furu Wei, and Ming Zhou. Neural latent extractive document summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 779–784. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/D18-1088>.

Xingxing Zhang, Furu Wei, and Ming Zhou. HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5059–5069, Florence, Italy, July 2019b. Association for Computational Linguistics. doi: 10.18653/v1/P19-1499. URL <https://www.aclweb.org/anthology/P19-1499>.