

Stepik. Neural networks and NLP. Список литературы.

Примечание: список литературы от Романа Суворова (из комментариев к степам)

Модуль 1. Введение

1.4 В общих чертах: Лингвистический анализ

Графематика <https://en.wikipedia.org/wiki/Graphemics>

Регулярные выражения:

- https://ru.wikipedia.org/wiki/%D0%A0%D0%B5%D0%B3%D1%83%D0%BB%D1%8F%D1%80%D0%BD%D1%8B%D0%B5_%D0%B2%D1%8B%D1%80%D0%B0%D0%B6%D0%B5%D0%BD%D0%B8%D1%8F
- <https://regexr.com/>

Токенизация - SpaCy Tokenizer <https://spacy.io/usage/spacy-101/>

Морфологический анализатор для русского языка PyMorphy2 <https://pymorphology2.readthedocs.io/en/latest/>

Список работ с кодом по морфологическому анализу <https://paperswithcode.com/task/morphological-analysis>

Как эффективно хранить словари -

Trie https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%B5%D1%84%D0%B8%D0%BA%D1%81%D0%BD%D0%BE%D0%B5_%D0%B4%D0%B5%D1%80%D0%B5%D0%B2%D0%BE

Разрешение частеречной омонимии - POS-теггинг (part of speech):

- TreeTagger - система, основанная на правилах <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

- POS-теггинг в SpaCy <https://spacy.io/usage/linguistic-features/#pos-tagging>
- POS-теггинг в Stanford NLP <https://nlp.stanford.edu/software/tagger.html>

Вероятностные модели, в том числе вероятностные модели последовательностей:

- Burr Settles, Probabilistic Sequence Models, <http://pages.cs.wisc.edu/~bsettles/ibs08/lectures/03-sequencemodels.pdf>
- Hidden Markov Models, HMM, https://en.wikipedia.org/wiki/Hidden_Markov_model
- Probabilistic Graphical Models, Coursera, <https://www.coursera.org/specializations/probabilistic-graphical-models>

Извлечение именованных сущностей:

- https://en.wikipedia.org/wiki/Named-entity_recognition
- AllenNLP демо с извлечением именованных сущностей <https://demo.allennlp.org/named-entity-recognition>
- DeepPavlov демо с извлечением именованных сущностей <https://demo.deepdpavlov.ai/#/en/ner>
- Томита-парсер - инструмент для извлечения сущностей и фактов с помощью правил <https://yandex.ru/dev/tomita/>
- Natasha - инструмент для извлечения сущностей и фактов с помощью правил <https://github.com/natasha/natasha>
- Список статей с кодом <https://paperswithcode.com/task/named-entity-recognition-ner>

Алгоритмы синтаксического разбора и анализаторы сдвиг-свёртка:

- https://en.wikipedia.org/wiki/Canonical_LR_parser
- Economopoulos, Giorgios Robert. *Generalised LR parsing algorithms*. Diss. University of London, 2006. <https://pdfs.semanticscholar.org/2ff7/4ea9a0147318bc19e30c6d0f72e29a5f92c3.pdf>
- Breveglieri, Luca, Stefano Crespi Reghizzi, and Angelo Morzenti. "Parsing methods streamlined." *arXiv preprint arXiv:1309.7584* (2013). <https://arxiv.org/abs/1309.7584>
- SyntaxNet <https://ai.googleblog.com/2016/05/announcing-syntaxnet-worlds-most.html>
- SyntaxNet TensorFlow <https://github.com/tensorflow/models/tree/master/research/syntaxnet>
- Nivre, Joakim, Johan Hall, and Jens Nilsson. "Maltparser: A data-driven parser-generator for dependency parsing." *LREC*. Vol. 6. 2006. http://lrec-conf.org/proceedings/lrec2006/pdf/162_pdf.pdf
- Список статей с кодом <https://paperswithcode.com/task/dependency-parsing>

Семантический анализ:

- https://en.wikipedia.org/wiki/Semantic_role_labeling.
- AllenNLP демо с семантическим анализом <https://demo.allennlp.org/semantic-role-labeling>.
- Yih, Scott Wen-tau, and Kristina Toutanova. "Automatic semantic role labeling." *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Tutorial Abstracts*. Association for Computational Linguistics, 2006.
- Gildea, Daniel, and Daniel Jurafsky. "Automatic labeling of semantic roles." *Computational linguistics* 28.3 (2002): 245-288.
- Neural Methods for Semantic Role Labeling <https://diegma.github.io/slides/TutorialNNforSRL.pdf>
- He, Luheng, et al. "Deep semantic role labeling: What works and what's next." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.
- Shelmanov A. O., Devyatkin D. A. Semantic Role Labeling with Neural Networks for Texts in Russian <http://www.dialog-21.ru/media/3945/shelmanovaodevyatkinda.pdf>
- Список работ с кодом <https://paperswithcode.com/task/semantic-role-labeling>

Извлечение отношений между сущностями:

- AllenNLP демо с извлечением сущностей в открытом домене и отношений между ними <https://demo.allennlp.org/open-information-extraction/>
- Fundel, Katrin, Robert Küffner, and Ralf Zimmer. "RelEx—Relation extraction using dependency parse trees." *Bioinformatics* 23.3 (2006): 365-371.
- Список ключевых статей и оценки качества http://nlpprogress.com/english/relationship_extraction.html
- Список статей по извлечению отношений (в основном нейроети) <https://github.com/roomylee/awesome-relation-extraction>
- Список репозиторий с реализациями извлечения отношений (в основном нейроети) <https://paperswithcode.com/task/relation-extraction>

Разрешение анафорических связей:

- [https://en.wikipedia.org/wiki/Anaphora_\(linguistics\)](https://en.wikipedia.org/wiki/Anaphora_(linguistics)).
- AllenNLP демо с разрешением кореференции <https://demo.allennlp.org/coreference-resolution/>
- Poesio, Massimo, Roland Stuckardt, and Yannick Versley. *Anaphora resolution*. Springer, 2016.
- Список работ с кодом <https://paperswithcode.com/sota/coreference-resolution-on-ontonotes>

Анализ дискурса:

- Shelmanov, Artem, et al. "Towards the Data-driven System for Rhetorical Parsing of Russian Texts." *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*. 2019.
- Yangfeng Ji, Jacob Eisenstein, Representation Learning for Text-level Discourse Parsing, Proceedings of ACL, 2014 <https://github.com/jiyfeng/RSTParser>
- Morey, Mathieu, Philippe Muller, and Nicholas Asher. "How much progress have we made on RST discourse parsing? A replication study of recent results on the RST-DT." 2017. <https://www.aclweb.org/anthology/D17-1136.pdf>

1.5 В общих чертах: Извлечение признаков

Извлечение признаков (общие сведения и метод - двоичный вектор, описывающий встречаемость слов в документе):

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- Извлечение признаков в Scikit-Learn https://scikit-learn.org/stable/modules/feature_extraction.html
- Извлечение признаков без разметки https://en.wikipedia.org/wiki/Feature_learning
- Линейные модели https://en.wikipedia.org/wiki/Linear_model
- Переобучение <https://en.wikipedia.org/wiki/Overfitting>

Извлечение признаков (метод - вектор вещественных чисел, описывающий встречаемость слов в документе, с учетом их частотности):

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- TF-IDF <https://ru.wikipedia.org/wiki/TF-IDF>
- Закон Ципфа https://ru.wikipedia.org/wiki/%D0%97%D0%B0%D0%BA%D0%BE%D0%BD_%D0%A6%D0%B8%D0%BF%D1%84%D0%B0

Извлечение признаков (метод - N-граммы символов и токенов, словосочетания):

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- N-граммы <https://ru.wikipedia.org/wiki/N-%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%B0>
- Разложение матриц https://en.wikipedia.org/wiki/Matrix_decomposition

Извлечение признаков (метод - Плотные векторные представления слов, предложений и текстов (word embeddings, doc embeddings):

- Дистрибутивно-семантические модели для русского языка <https://rusvectors.org/ru/>
- Предобученные языковые модели для нескольких языков <http://docs.deeppavlov.ai/en/master/features/models/bert.html>
- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book>

Извлечение признаков (метод - Ядерные методы (kernel methods) и графовые ядра):

- Закон Парето или правило 80/20 https://ru.wikipedia.org/wiki/%D0%97%D0%B0%D0%BA%D0%BE%D0%BD_%D0%9F%D0%B0%D1%80%D0%B5%D1%82%D0%BE
- Ядерные методы https://en.wikipedia.org/wiki/Kernel_method
- SVM с нелинейным ядром <https://towardsdatascience.com/understanding-support-vector-machine-part-2-kernel-trick-merciers-theorem-e1e6848c6c4d>
- Kriege, Nils M., Fredrik D. Johansson, and Christopher Morris. "A Survey on Graph Kernels." *arXiv preprint arXiv:1903.11835* (2019). <https://arxiv.org/abs/1903.11835>

1.6 Прикладные задачи обработки текста (которые так или иначе можно заложить в основу продукта)

Прикладные задачи обработки текста (классификация):

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- https://en.wikipedia.org/wiki/Bag-of-words_model
- <https://ru.wikipedia.org/wiki/TF-IDF>
- Линейные модели, логистическая регрессия https://en.wikipedia.org/wiki/Logistic_regression
- Линейные модели, SVM https://en.wikipedia.org/wiki/Support-vector_machine
- Обобщённые линейные модели https://en.wikipedia.org/wiki/Generalized_linear_model
- Список статей с кодом по тематике классификации текстов <https://paperswithcode.com/task/text-classification>
- Лексико-синтаксические шаблоны в системе GATE <https://gate.ac.uk/sale/tao/splitch8.html#chap:jape>

- Лексико-синтаксические шаблоны Томита-парсера <https://github.com/yandex/tomita-parser/blob/master/docs/ru/tutorial/basic-rules.md>

Прикладные задачи обработки текста (поиск: по запросу, по изображению и изображений по текстам, похожих текстов):

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- Feng, Yansong, and Mirella Lapata. "Automatic image annotation using auxiliary text information." *Proceedings of ACL-08: HLT*. 2008.
- Wang, Liwei, Yin Li, and Svetlana Lazebnik. "Learning deep structure-preserving image-text embeddings." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- Frome, Andrea, et al. "Devise: A deep visual-semantic embedding model." *Advances in neural information processing systems*. 2013.
- Gomaa, Wael H., and Aly A. Fahmy. "A survey of text similarity approaches." *International Journal of Computer Applications* 68.13 (2013): 13-18.
- Ianina, Anastasia, Lev Golitsyn, and Konstantin Vorontsov. "Multi-objective topic modeling for exploratory search in tech news." *Conference on Artificial Intelligence and Natural Language*. Springer, Cham, 2017.
- Zubarev, Denis, and Ilya Sochenkov. "Using Sentence Similarity Measure for Plagiarism Source Retrieval." *CLEF (Working Notes)*. 2014.

Прикладные задачи обработки текста (поиск: вопросно-ответный поиск):

- Chen, Danqi, et al. "Reading Wikipedia to Answer Open-Domain Questions." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.
- Rajpurkar, Pranav, et al. "SQuAD: 100,000+ Questions for Machine Comprehension of Text." *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2016.
- Ferrucci, David, et al. "Building Watson: An overview of the DeepQA project." *AI magazine* 31.3 (2010): 59-79.
- Gliozzo, Alfio, et al. "Semantic technologies in IBM watson." *Proceedings of the fourth workshop on teaching NLP and CL*. 2013.
- Shelmanov, A. O., et al. "Semantic-syntactic analysis for question answering and definition extraction." *Scientific and Technical Information Processing* 44.6 (2017): 412-423.

Прикладные задачи обработки текста (извлечение структурированной информации):

- Kelly, Liadh, et al. "Overview of the share/clef ehealth evaluation lab 2014." *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, Cham, 2014.

- Surdeanu, Mihai, et al. "Stanford's distantly-supervised slot-filling system." (2011).
- Peng, Fuchun, and Andrew McCallum. "Information extraction from research papers using conditional random fields." *Information processing & management* 42.4 (2006): 963-979.
- Mann, Gideon S., and David Yarowsky. "Multi-field information extraction and cross-document fusion." *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2005.
- Ratner, Alexander J., et al. "Snorkel: Fast training set generation for information extraction." *Proceedings of the 2017 ACM International Conference on Management of Data*. ACM, 2017.
- Dernoncourt, Franck, Ji Young Lee, and Peter Szolovits. "NeuroNER: an easy-to-use program for named-entity recognition based on neural networks." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 2017.
- Dalvi, Bhavana, et al. "IKE-an interactive tool for knowledge extraction." *Proceedings of the 5th Workshop on Automated Knowledge Base Construction*. 2016.
- Banko, Michele, et al. "Open information extraction from the web." *Ijcai*. Vol. 7. 2007
- Huang, Zhiheng, Wei Xu, and Kai Yu. "Bidirectional LSTM-CRF models for sequence tagging." *arXiv preprint arXiv:1508.01991* (2015).
- Arkhipov, Mikhail Y., and Mikhail S. Burtsev. "Application of a hybrid Bi-LSTM-CRF Model to the task of Russian named entity recognition." *Conference on Artificial Intelligence and Natural Language*. Springer, Cham, 2017.

Прикладные задачи обработки текста (диалоговые системы/чат-боты, машинный перевод):

- Chen, Hongshen, et al. "A survey on dialogue systems: Recent advances and new frontiers." *Acm Sigkdd Explorations Newsletter* 19.2 (2017): 25-35.
- Serban, Iulian Vlad, et al. "A survey of available corpora for building data-driven dialogue systems: The journal version." *Dialogue & Discourse* 9.1 (2018): 1-49.
- Serban, Iulian V., et al. "Building end-to-end dialogue systems using generative hierarchical neural network models." *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
- Chuklin, Aleksandr, et al. "Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI." *Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI*. 2018.
- Dinan, Emily, et al. "The second conversational intelligence challenge (convai2)." *arXiv preprint arXiv:1902.00098* (2019).
- Lowe, Ryan Thomas, et al. "Training end-to-end dialogue systems with the ubuntu dialogue corpus." *Dialogue & Discourse* 8.1 (2017): 31-65.
- Burtsev, Mikhail, et al. "DeepPavlov: open-source library for dialogue systems." *Proceedings of ACL 2018, System Demonstrations*. 2018.

- Gupta, Arshit, John Hewitt, and Katrin Kirchhoff. "Simple, Fast, Accurate Intent Classification and Slot Labeling." *arXiv preprint arXiv:1903.08268* (2019).

Прикладные задачи обработки текста (эксплоративный анализ):

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- O'Connor, Brendan, Michel Krieger, and David Ahn. "Tweetmotif: Exploratory search and topic summarization for twitter." *Fourth International AAAI Conference on Weblogs and Social Media*. 2010.
- Соченков, Илья Владимирович, Денис Владимирович Зубарев, and Илья Александрович Тихомиров. "Эксплоративный патентный поиск." *Информатика и её применения* 12.1 (2018): 89-94.
- Medlar, Alan, et al. "Pulp: A system for exploratory search of scientific literature." *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2016.
- Ianina, Anastasia, Lev Golitsyn, and Konstantin Vorontsov. "Multi-objective topic modeling for exploratory search in tech news." *Conference on Artificial Intelligence and Natural Language*. Springer, Cham, 2017.

Модуль 2. Векторная модель текста и классификация длинных текстов

2.1 Векторная модель текста и TF-IDF

Разреженные векторные модели (методы мешка слов), TF-IDF:

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- Количество уникальных слов в документе - закон Хипса https://ru.wikipedia.org/wiki/%D0%97%D0%B0%D0%BA%D0%BE%D0%BD_%D0%A5%D0%B8%D0%BF%D1%81%D0%B0
- Векторизация текстов через подсчёт количества словоупотреблений https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html
- Закон Ципфа https://en.wikipedia.org/wiki/Zipf%27s_law

Разреженные векторные модели (методы мешка слов), PMI - Pointwise mutual information (метод взаимной информации):

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <https://nlp.stanford.edu/IR-book/>
- Маргинальное (частное, маржинальное) распределение вероятностей https://en.wikipedia.org/wiki/Marginal_distribution
- Точечная взаимная информация https://en.wikipedia.org/wiki/Pointwise_mutual_information
- Применение PMI для представления смыслов слов (про это будут ещё лекции 3.2 и 3.3) Levy, Omer, and Yoav Goldberg. "Neural word embedding as implicit matrix factorization." *Advances in neural information processing systems*. 2014. <https://papers.nips.cc/paper/5477-neural-word-embedding-as-implicit-matrix-factorization.pdf>
- Взаимная информация - мера связанности двух случайных величин - мат.ожидание PMI https://en.wikipedia.org/wiki/Mutual_information
- Ещё несколько способов взвешивания и отбора признаков https://scikit-learn.org/stable/modules/feature_selection.html

2.2 Создаём нейросеть для работы с текстом

Линейные модели:

- Линейная модель https://en.wikipedia.org/wiki/Linear_classifier
- Логистическая регрессия в Scikit-learn https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- Обобщённые линейные модели https://en.wikipedia.org/wiki/Generalized_linear_model
- Метод наибольшего правдоподобия - откуда можно вывести формулу BCE https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

Чтобы усилить линейную модель, можно

- нагенерировать более мощных признаков, например, из комбинаций исходных признаков <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html>
- или использовать kernel trick - перейти из пространства признаков в пространство расстояний до других объектов https://scikit-learn.org/stable/modules/kernel_approximation.html

Под "более быстрыми методами оптимизации" имеются в виду приближённые методы второго порядка (квази-Ньютоновские):

- https://en.wikipedia.org/wiki/Limited-memory_BFGS
- <https://github.com/tensorflow/kfac>

Модуль 3. Базовые нейросетевые методы работы с текстами

3.3 Семинар: рецепты еды и Word2Vec на PyTorch

<https://rusvectors.org/ru/> - эмбединги для русского языка

<https://wikipedia2vec.github.io/wikipedia2vec/pretrained/> - претренированные эмбединги для разных языков

<https://wikipedia2vec.github.io/demo/> - очень интересная 3D визуализация эмбедингов (соседей)

3.6 Свёрточные нейросети для обработки текстов

Майкрософт завёз крутой сборник решений разных задач по NLP <https://github.com/microsoft/nlp-recipes>. Можно получить очень много опыта от изучения API и исходников. (Роман Суворов)

Модуль 4. Языковые модели и генерация текста

4.4 Агрегация, механизм внимания / 4.5 Трансформер и self-attention

От авторов курса и от студентов:

<https://arxiv.org/abs/1706.03762> (<https://arxiv.org/pdf/1706.03762.pdf>) - классика по механизму внимания (Vaswani A. et al. Attention is all you need //Advances in neural information processing systems. – 2017. – С. 5998-6008)

<https://habr.com/ru/post/341240/> - Transformer — новая архитектура нейросетей для работы с последовательностями

[Neural Machine Translation by Jointly Learning to Align and Translate, 2014](#)

<https://towardsdatascience.com/attention-based-neural-machine-translation-b5d129742e2c> - хорошая статья на TDS

<http://nlp.seas.harvard.edu/2018/04/03/attention.html> — очень хорошая аннотация оригинальной статьи Attention Is All You Need

<http://jalammar.github.io/illustrated-transformer/> - божественная визуализация алгоритмов трансформера для закрепления.

<https://towardsdatascience.com/attention-and-its-different-forms-7fc3674d14dc> - Attention and its Different Forms
https://kazemnejad.com/blog/transformer_architecture_positional_encoding/ - Transformer Architecture: The Positional Encoding

Модуль 5. Преобразование последовательностей: 1-к-1, N-к-M

5.4 Преобразование последовательностей (seq2seq)

Ключевые статьи по теме (от Романа Суворова):

- **Классический Seq2Seq:** Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems*. 2014. <https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf>
- **Seq2Seq с механизмом внимания:** Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." *arXiv preprint arXiv:1409.0473* (2014). <https://arxiv.org/pdf/1409.0473.pdf>
- **Google Neural Machine Translation:** Wu, Yonghui, et al. "Google's neural machine translation system: Bridging the gap between human and machine translation." *arXiv preprint arXiv:1609.08144* (2016). <https://arxiv.org/pdf/1609.08144.pdf>
- **Свёрточный Seq2Seq:** Gehring, Jonas, et al. "Convolutional sequence to sequence learning." *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017. <https://arxiv.org/pdf/1705.03122.pdf>
- **Transformer:** Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems*. 2017. <https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>
- **Какой энкодер и декодер лучше?** Chen, Mia Xu, et al. "The best of both worlds: Combining recent advances in neural machine translation." *arXiv preprint arXiv:1804.09849* (2018). <https://arxiv.org/abs/1804.09849>

Кроме того, из сносок к видео:

[1] [Статистическая теория обучения](#)

[2] [Recurrent Neural Networks \(RNN\) - The Vanishing Gradient Problem](#)

[3] Bengio Y., Simard P., Frasconi P. [Learning long-term dependencies with gradient descent is difficult](#) //IEEE transactions on neural networks. – 1994. – Т. 5. – №. 2. – С. 157-166.

[4] Pascanu R., Mikolov T., Bengio Y. [On the difficulty of training recurrent neural networks](#) //International conference on machine learning. – 2013. – С. 1310-1318.

- [5] Luong M. T., Pham H., Manning C. D. Effective approaches to attention-based neural machine translation // [arXiv preprint arXiv:1508.04025](#). – 2015.
- [6] Schuster M., Paliwal K. K. [Bidirectional recurrent neural networks](#) //IEEE transactions on Signal Processing. – 1997. – Т. 45. – №. 11. – С. 2673-2681.
- [7] Wu Y. et al. Google's neural machine translation system: Bridging the gap between human and machine translation // [arXiv preprint arXiv:1609.08144](#). – 2016.
- [8] Johnson M. et al. [Google's multilingual neural machine translation system: Enabling zero-shot translation](#) //Transactions of the Association for Computational Linguistics. – 2017. – Т. 5. – С. 339-351.
- [9] <https://ru.wikipedia.org/wiki/GPGPU>

- [1] Gehring J. et al. [Convolutional sequence to sequence learning](#) //Proceedings of the 34th International Conference on Machine Learning- Volume 70. – JMLR. org, 2017. – С. 1243-1252.
- [2] Gehring J. et al. [A convolutional encoder model for neural machine translation](#) //arXiv preprint arXiv:1611.02344. – 2016.
- [3] Vaswani, Ashish, et al. [Attention is all you need](#). Advances in neural information processing systems. 2017.

- [1] Chen, Mia Xu, et al. [The best of both worlds: Combining recent advances in neural machine translation](#). arXiv preprint arXiv:1804.09849 (2018).
- [2] Жадное декодирование и декодирование через лучевой поиск (beam-search) рассматривалось [в семинаре про трансформер](#).

Модуль 6. Transfer learning, адаптация моделей

- [1] [NLP's ImageNet moment has arrived](#), 12 July 2018

- [1] Peters M. E. et al. [Deep contextualized word representations](#) //arXiv preprint arXiv:1802.05365. – 2018.
- [2] Akbik A., Blythe D., Vollgraf R. [Contextual string embeddings for sequence labeling](#) //Proceedings of the 27th International Conference on Computational Linguistics. – 2018. – С. 1638-1649.
- [3] Baevski A. et al. [Cloze-driven pretraining of self-attention networks](#) //arXiv preprint arXiv:1903.07785. – 2019.
- [4] [The State of Transfer Learning in NLP](#), 18 August 2019

- [1] Devlin J. et al. [Bert: Pre-training of deep bidirectional transformers for language understanding](#) //arXiv preprint arXiv:1810.04805. – 2018.
- [1] Mikolov T. et al. [Distributed representations of words and phrases and their compositionality](#) //Advances in neural information processing systems. – 2013. – С. 3111-3119.
- [2] [Transformer-XL: Unleashing the Potential of Attention Models](#), January 29, 2019, Zhilin Yang and Quoc Le, Google AI
- [3] Dai Z. et al. [Transformer-xl: Attentive language models beyond a fixed-length context](#) //arXiv preprint arXiv:1901.02860. – 2019.
- [4] Young T. et al. [Recent trends in deep learning based natural language processing](#) //IEEE Computational Intelligence Magazine. – 2018. – Т. 13. – №. 3. – С. 55-75.
- [5] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. [Neural machine translation by jointly learning to align and translate](#). arXiv preprint arXiv:1409.0473 (2014).

GPT-2:

- [1] <https://openai.com/blog/unsupervised-sentiment-neuron/>
- [2] <https://github.com/eukaryote31/openwebtext>
- [3] <https://paperswithcode.com/task/reading-comprehension>
- [4] <https://paperswithcode.com/task/zero-shot-learning>

Из комментариев (студента):

В качестве дополнения (вдруг кому-то понадобится или просто интересно):

Russian GPT-2 https://github.com/mgrankin/ru_transformers

демо "Порфирьевич" <https://porfirevich.ru/>

Статьи первоисточники, на основе которых подготовлены видеолекции степа (рекомендованы к прочтению):

- [1] <http://jalamar.github.io/illustrated-transformer/>
- [2] <http://jalamar.github.io/illustrated-bert/>
- [3] <http://web.stanford.edu/class/cs224n/slides/cs224n-2020-lecture14-contextual-representations.pdf>
- [4] <https://medium.com/mlreview/understanding-building-blocks-of-ulmfit-818d3775325b>

- [1] Корпус коротких текстов Юлии Рубцовой: <http://study.mokoron.com/>

[2] Liu Y. et al. [Roberta: A robustly optimized bert pretraining approach](#) //arXiv preprint arXiv:1907.11692. – 2019.

[2] <https://rajpurkar.github.io/SQuAD-explorer/> - датасет SQUAD для вопросно-ответного поиска (с топом лучших моделей)

Что еще почитать

Список ссылок на статьи про архитектуры, рассмотренные в лекции:

1. OpenAI Transformer: [Improving Language Understanding by Generative Pre-Training](#)
2. ELMO: [Deep contextualized word representations](#)
3. BERT: [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#)
4. Transformer-XL: [Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context](#)
5. GPT-2: [Language Models are Unsupervised Multitask Learners](#)

Серия статей от [Jay Alammar](#) (отличные посты с очень понятными картинками):

1. [Illustrated Transformer](#)
2. [Illustrated BERT](#) (материалы отсюда были использованы в лекции)
3. [Illustrated GPT-2](#)

Еще немного блог-постов:

1. [ELMO от AllenNLP](#)
2. [Transformer-XL от Google AI](#)
3. [Простая статья на Medium про Transformer-XL](#)

Ссылки на код моделей:

1. [OpenAI Transformer](#)
2. [ELMO \(Tensorflow\)](#)
3. [BERT \(Tensorflow\)](#)
4. [Transformer-XL](#)
5. [GPT-2](#)
6. Обновление от того же автора по GPT-3: <http://jalamar.github.io/how-gpt3-works-visualizations-animations/>

Модуль 7. Заключение

Рекомендуемые курсы по NLP и смежным направлениям:

1. Stanford CS224 [страница курса](#)
2. Coursera: ["Обработка текстов, написанных на естественных языках"](#)
3. Coursera: ["Графические вероятностные модели"](#)
4. Fast.ai: [YouTube](#), [блог-пост](#), [репозиторий github](#)
5. Microsoft edX: ["Natural Language Processing"](#)
6. Udacity: ["Become a Natural Language Processing Expert"](#)
7. Stepik: ["Введение в обработку естественного языка"](#)

Инструменты:

1. [rusvectors](#): семантические модели для русского языка
2. [youtokentome](#): инструмент для быстрой токенизации текста от команды VK
3. [Список популярных датасетов](#)
4. [Вики ACL Anthology](#)
5. Вики ACL: [раздел с задачами](#)

Статьи:

- [scholar.google.com](#)
- [arxiv.org](#)
- [arxiv-sanity.org](#)
- [paperswithcode.com](#)

Блоги ученых:

- [Ruder.io](#)

Блоги компаний:

- [OpenAI блог](#)
- [Google AI блог](#)

- [Facebook AI Research \(FAIR\) блог](#)

Другие интересные ресурсы:

- [NL Pub](#)
- [NeuroNuggets](#)
- [distill.pub](#)

Очень крутой сборник ресурсов по NLP: [NLP Awesome List](#)

Облачные сервисы:

[Github Student Pack](#)

[Microsoft Azure для учащихся](#)

[AWS Educate](#)