

7.1 – Как развиваться и что читать, чтобы стать экспертом в NLP?

Модуль 7 - Финальное соревнование на Kaggle и заключение

Анастасия Янина

Samsung AI Center Moscow

Совет №1: смотрите онлайн-курсы

- Stanford CS224: [страница курса](#)



CS224n: Natural Language Processing with Deep Learning



YouTube UA Введите запрос

Lecture Collection | Natural Language Processing with Deep Learning (Winter 2017)

19 видео • 796 594 просмотра • Обновлено 3 апр. 2017 г.

Natural language processing (NLP) deals with the key artificial intelligence technology of understanding complex human language communication. This lecture series provides a thorough introduction to the cutting-edge research in deep learning applied to NLP, an approach that has recently obtained very high performance across many different NLP tasks including question answering and machine translation.

- Lecture 1 | Natural Language Processing with Deep Learning**
Stanford University School of Engineering
1:11:41
- Lecture 2 | Word Vector Representations: word2vec**
Stanford University School of Engineering
1:18:17
- Lecture 3 | GloVe: Global Vectors for Word Representation**
Stanford University School of Engineering
1:18:40
- Lecture 4: Word Window Classification and Neural Networks**
Stanford University School of Engineering
1:16:43
- Lecture 5: Backpropagation and Project Advice**
Stanford University School of Engineering
1:18:20
- Lecture 6: Dependency Parsing**
Stanford University School of Engineering
1:23:07

Совет №1: смотрите онлайн-курсы

- **Stanford CS224:** [страница курса](#)
- Coursera
 - [“Обработка текстов, написанных на естественных языках”](#)
(специализация “Продвинутое машинное обучение” на Coursera) от ВШЭ и Яндекса

4.6 ★★★★★

Рецензии: 116 >



38%

начал новую карьеру,
пройдя эти курсы



36%

получил значимые
преимущества в
карьере благодаря
этому курсу



33%

стал больше
зарабатывать или
получил повышение

Лучшие отзывы о курсе Обработка текстов, написанных на естественных языках

★★★★★ автор: GY • МАРТ 24-ГО 2018

Great thanks to this amazing course! I learned a lot on state-to-art natural language processing techniques! Really like your awesome programming assignments! See you HSE guys in next class!

★★★★★ автор: YY • ЯНВ. 2-ГО 2019

I like this course very much. It is a good introduction for NLP. But if you want to know more about the NLP, you need to search and read a lot of posts during the learning process.

Совет №1: смотрите онлайн-курсы

- **Stanford CS224:** [страница курса](#)
- Coursera
 - “[Обработка текстов, написанных на естественных языках](#)”
(специализация “Продвинутое машинное обучение” на Coursera) от ВШЭ и Яндекса
 - Специализация “[Графические вероятностные модели](#)”
- Курс от Fast.AI
 - Видео на [YouTube](#)
 - [Блог-пост](#)
 - [Github репозиторий](#)
- Курс от Microsoft на edX “[Natural Language Processing](#)”
- “[Become a Natural Language Processing Expert](#)” на Udacity (3 месяца)

Совет №2: изучайте код библиотек, имплементации моделей

- `pytorch >=1.2`
- `torchtext`
- `pytorch-transformers`
- `spacy`
- `fast.ai`
- `Gensim`
- `NLTK`
- `deeppavlov`
- `AllenNLP`

Совет №2: изучайте код библиотек, имплементации моделей

Еще полезные инструменты и ссылки

- [rusvectors](#): семантические модели для русского языка

RusVectōrēs: семантические модели для русского языка

'You shall know a word by the company it keeps.' (Firth 1957).

Введите слово, чтобы получить список из 10 его ближайших семантических ассоциатов (квази-синонимов).

Будет использована модель, обученная на Википедии и Национальном корпусе русского языка; другие модели вы можете найти на вкладке [Похожие слова](#).

Найти похожие слова!

Совет №2: изучайте код библиотек, имплементации моделей

Еще полезные инструменты и ссылки

- [rusvectors](#): семантические модели для русского языка
- [youtokentome](#): инструмент для быстрой токенизации текста от команды VK
-
- [Список популярных датасетов](#)
- [Вики ACL Anthology](#)
 - [Особенно раздел с задачами!](#)

Совет №3: следите за новыми статьями, но не тратьте на это все рабочее время :)

- scholar.google.com
- arxiv.org
- arxiv-sanity.org
- paperswithcode.com

[Browse](#) > Natural Language Processing

Natural Language Processing

📊 405 leaderboards • 222 tasks • 106 datasets • 3223 papers with code

Machine Translation



Machine
Translation

📊 43 leaderboards

521 papers with code



Transliteration

15 papers with code



Unsupervised
Machine
Translation

📊 9 leaderboards

9 papers with code



Multimodal
Machine
Translation

7 papers with code



Low-Resource
Neural
Machine
Translation

6 papers with code

► [See all 6 tasks](#)

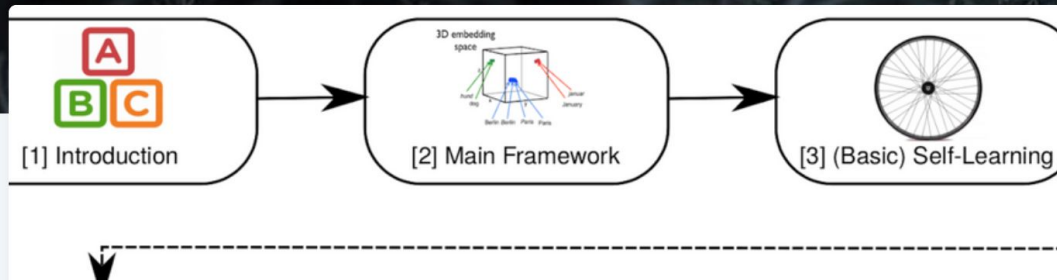
Совет №4: читайте блоги, обзоры на статьи

- [Ruder.io](https://sebastianruder.com/)

Sebastian Ruder

I'm a research scientist at DeepMind. I blog about Machine Learning, Deep Learning, and Natural Language Processing.

[ABOUT](#) [TAGS](#) [PAPERS](#) [TALKS](#) [NEWS](#) [FAQ](#) [SIGN UP FOR NLP NEWS](#) [NLP PROGRESS](#) [CONTACT](#)



CROSS-LINGUAL

Unsupervised Cross-lingual Representation Learning

Совет №4: читайте блоги, обзоры на статьи

- [Ruder.io](https://ruder.io)
- Блоги компаний:
 - [OpenAI блог](https://openai.com/blog)



[ABOUT](#) [PROGRESS](#) [RESOURCES](#) [BLOG](#)

Progress

OpenAI works on advancing AI capabilities, safety, and policy.

[RELEASES](#) [PAPERS](#)

Milestone Releases

Совет №4: читайте блоги, обзоры на статьи

- [Ruder.io](https://ruder.io)
- Блоги компаний:
 - [OpenAI блог](https://openai.com/blog)
 - [Google AI блог](https://ai.google.dev)



The latest news from Google AI

Совет №4: читайте блоги, обзоры на статьи

- [Ruder.io](https://ruder.io)
- Блоги компаний:
 - [OpenAI блог](#)
 - [Google AI блог](#)
 - [Facebook AI Research \(FAIR\) блог](#)
 - ...

facebook Artificial Intelligence

Совет №4: читайте блоги, обзоры на статьи

- [Ruder.io](https://ruder.io)
- Блоги компаний:
 - [OpenAI блог](https://openai.com/blog)
 - [Google AI блог](https://ai.google.dev)
 - [Facebook AI Research \(FAIR\) блог](https://research.facebook.com)
 - ...
- [NLPub](https://nlp.pub)
- [NeuroNuggets](https://neuronuggets.com)
- distill.pub

Очень крутой сборник ресурсов по NLP:

[NLP Awesome List](https://nlpawesome.org)

Совет №5: купите GPU или пользуйтесь облачными сервисами

- [Github Student Pack](#)
- [Microsoft Azure для учащихся](#)
- [AWS Educate](#)



Access to the AWS cloud, free training, and collaboration resources

Benefit Student Developer Pack members receive up to \$110 in bonus AWS credits for a total of \$75-\$150.



Simple cloud hosting, built for developers

Benefit \$50 in platform credit for new users.



Domain names, web hosting, and websites. Unicorns and rainbows come standard with our customer support.

Benefit One free domain name and free Advanced Security (SSL, privacy protection, and more).



Affordable registration, hosting, and domain management

Benefit 1 year SSL certificate.

Benefit 1 year domain name registration on the .me TLD.



Microsoft Azure is an ever-expanding set of cloud services to help you build, manage, and deploy applications.

Benefit Level up your ideas by learning cloud computing skills and get a \$100 credit with your free Azure for Students account—no credit card required.



Level up on trending coding skills at your own pace with interactive, text-based courses.

Benefit Get 6 free months of 60+ courses covering in-demand topics like Web Development, Python, Java, and Machine Learning.

Совет №6: участвуйте в конференциях, летних школах, хакатонах

Конференции

NLP

ACL
EMNLP
NAACL
EACL
COLING
CONLL
SIGIR
RANLP
LREC

ML

ICML
ICLR
AAAI
ICDM
IJCAI
NeurIPS

Соревнования

- Kaggle
- SemEVAL
- CLEF
- TREC
- Dialog
- RuEVAL

Совет №7: применяйте полученные знания на практике!

