

Слон и Аполлон

Генеративные трансформеры и поэзия

SBER

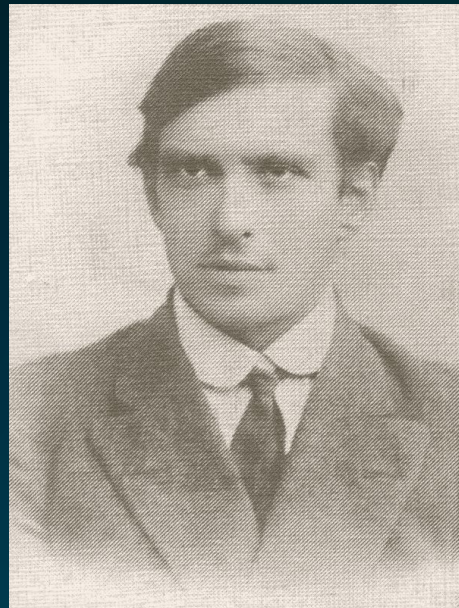
DEVICES

Алексей Константинович Лозина-Лозинский

* * *

Покорность Господу в слоне сильна, как видно,
(В зверинце я слона упорно наблюдал).
Он как профессор жил — разумно и солидно
И добрым хоботом под животом чесал.

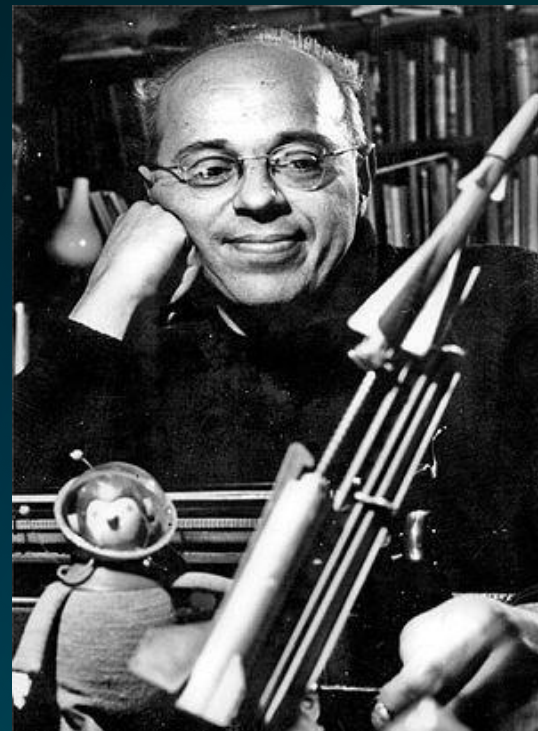
О, слон! Ведь ты мудрец. И, как мудрец, не смеешь
Ни сторожа побить, ни выйти из тюрьмы...
За это, тяжкий слон, когда ты околеешь,
Тебя к святым слонам должны причислить мы.

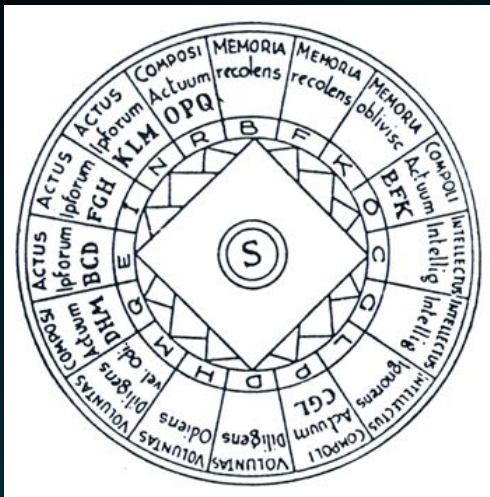


1914

«Поэты организовали несколько митингов протеста и потребовали опечатать машину, но никто, кроме них, не обращал внимания на феномен. Редакции газет были даже довольны, поскольку Электрувер, писавший под несколькими тысячами псевдонимов сразу, представлял готовую поэму заданных размеров на любой случай, и эта поэзия, хоть и на заказ, была такого качества, что читатели раскупали газеты нарасхват, а улицы так и пестрели лицами, полными неземного блаженства, мелькали бессознательные улыбки и слышались тихие всхлипывания. Стихи Электрувера знали все; воздух сотрясали хитроумнейшие рифмы, а наиболее впечатлительные натуры, потрясенные специально сконструированными метафорами или ассонансами, даже падали в обморок; но и к этому был подготовлен титан вдохновения: он сразу же вырабатывал соответствующее количество отрезвляющих сонетов».

Станислав Лем. Путешествие первое А, или Электрувер Трурля





В XIII веке каталонский миссионер, поэт, философ и теолог Раймунд Луллий создал «логическую машину», которая состояла из нескольких подвижных концентрических кругов, разделённых на сектора («камеры»), в которые были обозначены общие понятия или основные категории всего сущего. Поскольку каждый из кругов мог вращаться независимо от остальных, в результате вращения можно было добиться появления различных комбинаций понятий, в которых Луллий искал различные новые смыслы. Изобретатель утверждал, что конструкция машины была открыта ему свыше, в видении, посетившим его на родном острове Майорка



«СЧцГ кБяпэшиШыЫЗё.ЗьмыйВЦДЩ!К,яТ-ЪоСФжВТ-О» — как тебе такое, Илон Маск?

Николай Чуковский (Николай Корнеевич Чуковский, 1904—1965), один из учеников Николая Гумилёва (Николай Степанович Гумилёв, 1886—1921), вспоминал, что Гумилёв вёл семинар, на котором его участники писали стихи при помощи составленных мэтром таблиц, которые, по мнению Гумилёва, давали возможность сочинять стихи «без риска написать плохо». Знакомый Сергея Есенина (Сергей Александрович Есенин, 1895—1925) Матвей Ройзман (Матвей Давидович Ройзман, 1896—1973) вспоминал, что однажды, придя домой к поэту, он застал такую картину: *«Лежал Сергей на ковре, сбоку от него находилась небольшая старая коробочка от лото, а перед ним валялись нарезанные из карточек картонные квадратики: на одной стороне — цифры, на другой — написанные рукой Есенина слова. Он сказал, что пытается механизировать процесс сочинения стихов и образов. Взял из кучи несколько квадратиков, прочитал: Вечер, свечи, — и произнёс вслух:*

*Вдали розовый вечер
Зажёг жёлтые свечи...»*

«Остап простер свое любопытство вплоть до того, что ознакомился даже со статьей Ухудшанского «Улучшить работу лавочных комиссий». После этого Бендер еще оглядел диковинного журналиста с ног до головы, нехорошо улыбнулся и, почувствовав знакомое волнение стрелка-охотника, заперся в купе, Оттуда он вышел только через три часа, держа в руках большой разграфлённый, как ведомость, лист бумаги.

— Пишете? — вяло спросил Ухудшанский.

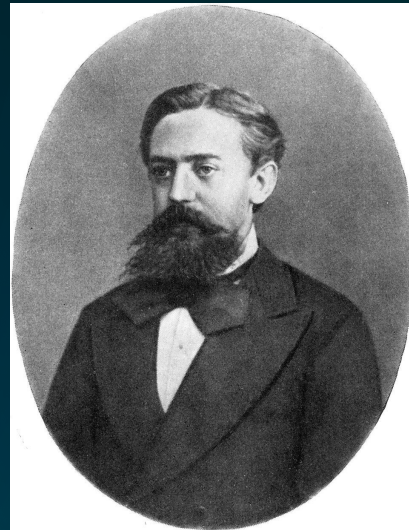
— Специально для вас, — ответил великий комбинатор. — Вы, я замечая, все время терзаетесь муками творчества. Писать, конечно, очень трудно. Я, как старый передовик и ваш собрат по перу, могу это засвидетельствовать. Но я изобрёл такую штуку, которая избавляет от необходимости ждать, покуда вас окатит потный вал вдохновения. Вот. Извольте посмотреть.

И Остап протянул Ухудшанскому лист, на котором было написано:

ТОРЖЕСТВЕННЫЙ КОМПЛЕКТ

НЕЗАМЕНИМОЕ ПОСОБИЕ ДЛЯ СОЧИНЕНИЯ ЮБИЛЕЙНЫХ СТАТЕЙ, ТАБЕЛЬНЫХ
ФЕЛЬЕТОНОВ, А ТАКЖЕ ПАРАДНЫХ СТИХОТВОРЕНИЙ. ОД И ТРОПАРЕЙ»

23 января 1913 года на заседании Физико-Математического Отделения Императорской Академии наук уже знакомый нам Андрей Андреевич Марков (старший) представил работу под названием «Пример статистического исследования над текстом «Евгения Онегина», иллюстрирующий связь испытаний в цепь», в которой исследовал статистические закономерности, обнаруживающиеся в 20 000 символах поэмы (первая глава и шестнадцать строф второй главы)



С августа 1953 года по май 1954 года на доске объявлений компьютерного факультета Манчестерского университета появлялись странные любовные письма:

ГОЛУБУШКА ЛЮБИМАЯ

*ВЫ МОЯ НЕЖНАЯ МЕЧТА. МОЯ ТОСКУЮЩАЯ СТРАСТЬ ПЫЛКО ВЗЫВАЕТ К ВАШЕМУ ЖЕЛАНИЮ.
МОЁ НЕОБУЗДАННОЕ ВЛЕЧЕНИЕ ТОСКУЕТ ПО ВАШЕЙ ПРЕДАННОСТИ. ВЫ МОЯ ОСТРАЯ
ПРИВЯЗАННОСТЬ: МОЯ СТРАСТНАЯ ЛЮБОВЬ.*

ВАШ НЕРАВНОДУШНЫЙ

M. U. C.

[DUCK DARLING

*YOU ARE MY FOND DESIRE. MY WISTFUL INFATUATION ARDENTLY ATTRACTS YOUR APPETITE. MY
LONGING YEARNs FOR YOUR DEVOTION. YOU ARE MY KEEN AFFECTION: MY PASSIONATE LOVE.*

YOURS CURIOUSLY

M. U. C.]

Аббревиатура MUC расшифровывалась как «Компьютер Манчестерского Университета» [Manchester University Computer]. Кристофер Стрейчи использовал встроенный генератор псевдослучайных чисел Ferranti Mark I и несколько списков слов для создания эмоциональных текстов. Благодаря усилиям программистов-реконструкторов, воспользоваться детищем Стрейчи можно и сегодня

Вот как описывает работу своей программы POETRY GENERATOR (позже — AUTOMATIC POETRY GENERATOR) её создательница Розмари Уэст (Rosemary K. West): *«Мой подход... основан на использовании богатого словаря слов и фраз, которые выбираются случайным образом, и из них формируются словосочетания согласно набору грамматических правил. Рассмотрим, например, следующее четверостишие: «Дерево погружает голые пальцы / в черное озеро льда / словно три серых гуся / сползают по снежному склону». Каждую строку стихотворения можно разбить на несколько частей... Дерево — подлежащее; опускает — сказуемое; голые пальцы — дополнение. Определив части, я могу вместо них подставлять другие слова из многочисленного (от 100 до 400 [штук]) набора для каждой части. Эти подстановки выбираются компьютером случайным образом. Например, воспользовавшись той же стихотворной структурой, я могла бы получить другое четверостишие: «Женищина прячет пять серых котят / под старым ржавым рыдваном / и тут же печальные клоуны / входят в ваш музей страданий».*

Любовница супруга моего

Любовница супруга моего как будто торжествует.

Но, глядя, как работаю я в храме,

В пещерах зла повсюду меня ищет,

И помнит обо мне, в постели засыпая.

Хотя живу я в царстве безразличья,

Не чувствует себя уверенно она.

И никогда в мечтах о власти винить меня она не перестанет.

как позвонить в сбербанк россии
возьму в аренду самосвал
лечение гипертонии
стендап пузатый генерал
измерить скорость интернета
литературная газета
фильм достучаться до небес
код магнитолы мерседес
вакансии в североморске
что значит слово реапод
несчастный случай снег идёт
цена на шубы в пятигорске
лес пахнет дубом и сосной
я просто должен быть с тобой

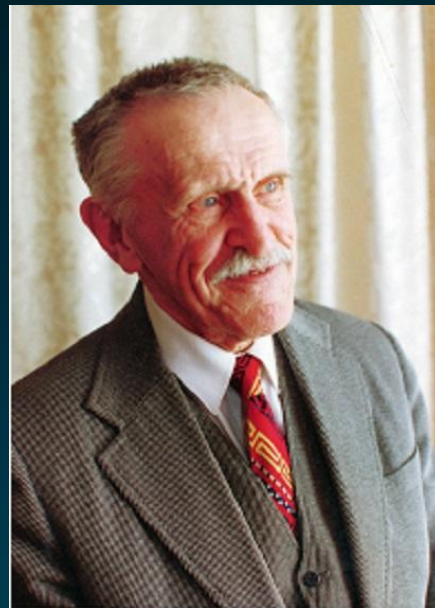
Яндекс.Автопоэт

Нид ворону

*Ври, вран —
вор ран,
гриф гроз,
гусь слез.
Сохл, плох,
трухл, рухл,
дряхл, чахл,
затхл, тухл.
Жухл, слаб
жрун жаб.
Дрянъ — вран,
дрянно, дран!*

Драпа мечу

*Меч, меч —
сери сеч,
бич битв,
брат бритв.
Быстр, остр,
шустр, востр,
страж страд,
страх стад.
Мчи, меч,
бей, меч —
битв бич,
сери сеч!*



*Ничего, ничего, кроме будничных дел
Я хотел бы уехать, куда не хотел
Я буду ходить по дорогам своим,
Но пока мы сидим и страдаем, скулим*

Припев:

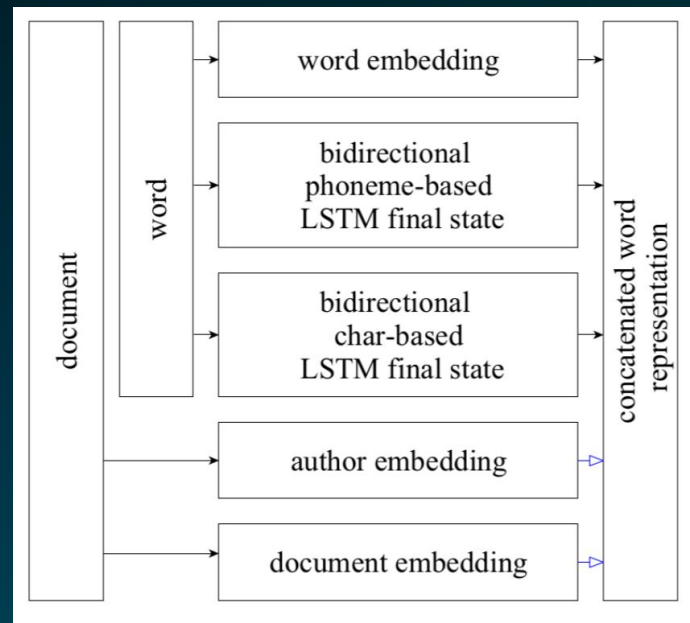
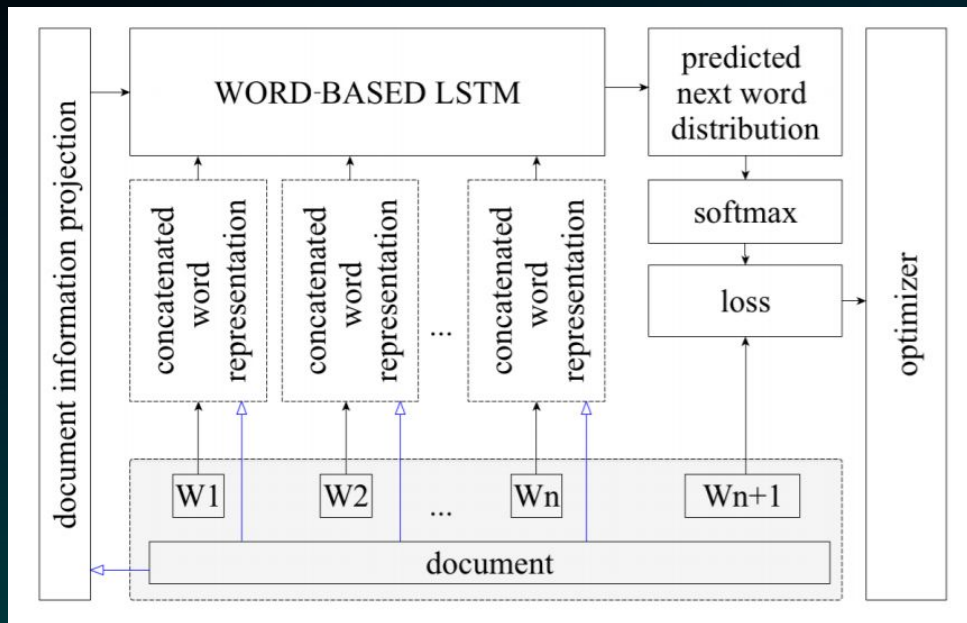
*Будет свет
И будет
Будет свет
Ничего
Будет свет
И будет
Будет вечный смех*

В 2016 году команда сотрудников «Яндекса» под руководством Алексея Тихонова и Ивана Ямщикова дала жизнь проекту под названием «Нейронная оборона». Нейросетевая генеративная модель «написала» тексты песен в стиле панк-рок-группы «Гражданская оборона» Егора Летова, а затем авторы записали на основе полученных текстов музыкальный альбом, стилизованный по звучанию под произведения «ГРОб».

Вот что пишет об этом проекте один из его авторов — Иван Ямщиков: *«После того, как Алексей Тихонов показал мне «нейропоэта», мы так сильно захотели использовать стихи, написанные нейросетью, что даже кушать не могли. Понятно было несколько вещей:*

- 1. Надо попробовать сделать стилизацию чего-нибудь культового,*
- 2. Надо, чтобы культовый оригинал был достаточно маргинален.*

Поясню, что имею ввиду под пунктом два. Хармс, к примеру, у нейросетки получался годный, мощный у неё выходил Хармс, а вот Александр Сергеевич Пушкин выходил сомнительный. Нейросеть хорошо стилизовала тексты авторов, которые предпочитали диссоциированный нарратив (скажем мягко) или вообще работали с эстетикой абсурда. В целом, можно было бы заморочиться, найти каких-нибудь актёров и заставить их читать нейрохармса на камеру (кстати, если заморочиться и найдёте, то я только за в это вписаться), но в этом не было какого-то достаточного художественного жеста, поэтому мы решили записать альбом. Сами, на коленке».



«Нейро-Пушкин»:

*во славу вакха или тьмы
мы гордо пировали*

«Нейро-Есенин»:

*ты под солнцем стоишь и в порфире
как в шелку беззаботно горишь*

«Нейро-Земфира»:

*ветер в голове
с красной тенью шепчется*

«Нейро-Летов»:

*только в ушах отражается даль
только белая смерть превращается в ад*

Для обучения своей модели Ямщиков и Тихонов использовали набор из 440 000 документов (110 000 на английском языке и 330 000 на русском), принадлежавших перу 20 700 авторов (19 000 для англоязычной части датасета и 1 700 для русскоязычной). Для тестирования качества генерации текста были выбраны тексты семи англоязычных и пяти русскоязычных авторов. В число первых вошли произведения Уильяма Шекспира, Эдгара По, Льюиса Кэрролла, Оскара Уайльда и Боба Марли, а также тексты песен музыкальных групп Nirvana и Muse. Для русского языка были отобраны произведения Пушкина, Есенина, Бродского, Егора Летова и Земфиры. Для того, чтобы оценить качество генерации текста, Тихонов и Ямщиков заставляли модель продолжать случайно выбранные фрагменты из произведений перечисленных выше авторов, отсутствующих в обучающей выборке. Затем результаты генерации сравнивались с оригинальными текстами при помощи метрики BLEU.

«Нейро-Пушкин»:

*во славу вакха или тьмы
мы гордо пировали*

«Нейро-Есенин»:

*ты под солнцем стоишь и в порфире
как в шелку беззаботно горишь*

«Нейро-Земфира»:

*ветер в голове
с красной тенью шепчется*

«Нейро-Летов»:

*только в ушах отражается даль
только белая смерть превращается в ад*

Для обучения своей модели Ямщиков и Тихонов использовали набор из 440 000 документов (110 000 на английском языке и 330 000 на русском), принадлежавших перу 20 700 авторов (19 000 для англоязычной части датасета и 1 700 для русскоязычной). Для тестирования качества генерации текста были выбраны тексты семи англоязычных и пяти русскоязычных авторов. В число первых вошли произведения Уильяма Шекспира, Эдгара По, Льюиса Кэрролла, Оскара Уайльда и Боба Марли, а также тексты песен музыкальных групп Nirvana и Muse. Для русского языка были отобраны произведения Пушкина, Есенина, Бродского, Егора Летова и Земфиры. Для того, чтобы оценить качество генерации текста, Тихонов и Ямщиков заставляли модель продолжать случайно выбранные фрагменты из произведений перечисленных выше авторов, отсутствующих в обучающей выборке. Затем результаты генерации сравнивались с оригинальными текстами при помощи метрики BLEU.

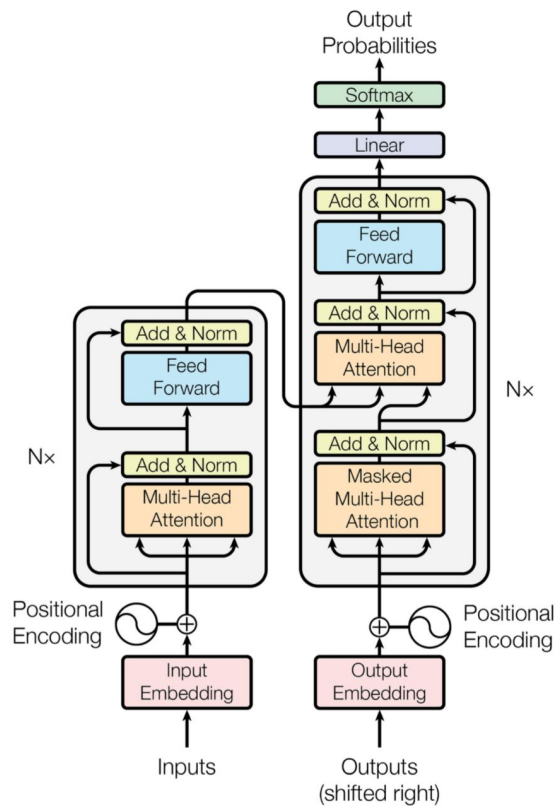
на фоне формы каменной границы,
как будто бы равнина далека,
лениво стынет полная столица,
чернеет свечка синяя строка.
а я спешу к победе бесконечной,
лечу с отрадой томной чередой,
но не томлюсь томительно сердечной
нежданно грустной памятью былой

[в лесу родилась ёлочка,
в лесу она росла,
зимой и летом стройная,
зелёная] была.
громада поднимается,
летает борода,
мелькает одиночество,
ложится пустота.
ликует революция,
шагает самолёт,
чернеет рана чистая,
синее небо свод.

...

лениво качается,
дымится столица,
клубится стремительно,
бушует зарница.
несётся колокольчик,
рыдает долина,
волнуется комната,
мечтает картина,
грустит королева,
болит голова,
гуляет корова,
хохочет сова.

Трансформеры



Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

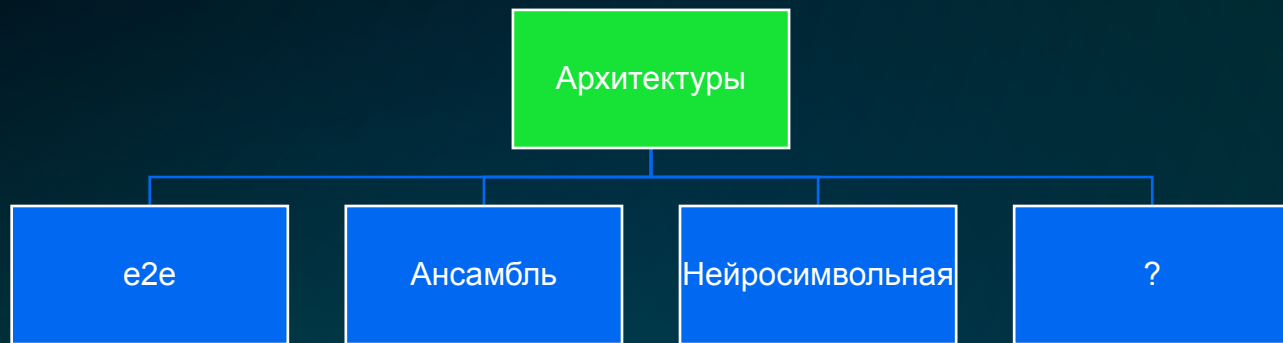
Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin*[‡]
illia.polosukhin@gmail.com

*Мир без царей и без власти
И без революций,
Забывтый, неожиданный и богатый,
Где вздымаются к небу
Языческие племена!
В полях, в лугах, на озёрах,
В лесах, в рощах, в полях,
Где ветер свищет и свистит,
Где в снегах радуга горит,
Где струны поют, звенят!
Где все, что нам дышит, дышит,
Как солнце и месяц единый,
Пелым живых, к нам приходит
С запада, с востока, с востока
На пёстром, праздничном троне
Славный царь Емельян.
И видна за ним страна*

*Избённая, изломленная
И дорогая...
Спит в ней гордый царь Емельян,
Спит народ и царь мирно спит,
Спит голова его алмазная,
Спит краса его стальная.
Спит краса его голубая,
Спит цвет его, краса его золотая,
Спит рабская честь его,
Спит вера его, и совесть его,
Спят и дремлют богатыри.*





Сэмплинг

```
graph TD; A[Сэмплинг] --- B[Nucleus sampling]; A --- C[Beam search]; A --- D[Pseudo-Gibbs]; A --- E[?];
```

Nucleus
sampling

Beam
search

Pseudo-G
ibbs

?



Сергей Марков

17 марта, 21:29 · 🌐



мОй| |дЯдЯ| |сАмЫх| |чЕстных| |прАвил|,
когДА| |нЕ| |в| |шУтку| |занемОг|,
Он| |уважАть| |себЯ| |застАвил
И| |лУчше| |выдумАть| |нЕ| |мОг|.
егО| |женА| |совсЕм| |больнАя|
нА| |свАдЬбе| |нЕ| |смоглА| |земнАя|
для| |брАтА| |мУжА| |быть| |егО|
хозяЙкой| |нА| |рукАх| |чегО
емУ| |совсЕм| |давнО| |бывАло|
чегО| |совсЕм| |когДА| |И| |прО|
чегО| |емУ| |сказАть| |нутрО|
чегО| |емУ| |совсЕм| |отвАлА|
когДА| |чегО| |совсЕм| |когДА|
чегО| |совсЕм| |когДА| |балДА|

Здравствуйте дорогой Мартин Алексеевич!



41

Комментарии: 28



Нравится



Комментировать



Поделиться

