**МИНОБРНАУКИ РОССИИ**

**Федеральное государственное бюджетное**

**образовательное учреждение высшего образования**

**«Тульский государственный университет»**

Подразделение: кафедра «Вычислительная техника»

**ОТЧЕТ**

|  |  |
| --- | --- |
| Вид практики | НИР (получение первичных навыков НИР) |
| Курс | 3 курс |
| Направление подготовки  /специальность | 09.03.01 Информатика и вычислительная техника |
| Ф.И.О. обучающегося | Шайхаттаров Дамир Владимирович |
| Место прохождения практики | ФГБОУ ВО «Тульский Государственный университет», каф. ВТ |
| Период прохождения практики | С 21 сентября по 17 октября 2020 г. |

Руководитель практики от

профильной организации (при наличии)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(Ф.И.О., должность) (подпись)*

М. П.

Руководитель практики от подразделения

Неелова Н.В., ст. преподаватель каф. ВТ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(Ф.И.О., должность) (подпись)*

г. Тула  
 2020 г.

РЕФЕРАТ

GOOGLE TRANSLATE, НЕЙРОСЕТЬ, ПЕРЕВОД, НЕЙРОННЫЕ СЕТИ, ПЕРЕВОД

Нужно понимать, что в данной НИР можно рассматривать несколько видов машинного перевода: SMT, RBMT, HMT и наконец NMT. Каждый вид имеет свои особенности и свои преимущества отчего всё ещё активно используется различными компании для выполнения своих задач, поэтому NMT – это не универсальный способ машинного перевода, но мне была интересна компания Google, а она в 2016 году решила перейти уйти от SMT к NMT, создав свою систему нейронного машинного перевода GNMT для проекта Google Translate. И исходя из того, что Google – одна из передовых компаний в сфере разработки, а также нейронные сети являются достаточно новой и интересной технологией мы возьмём объектом исследования GNMT, применяемой в Google Translate.

Цель работы - получение навыков выполнения НИР по проектированию систем машинного перевода, работы с глобальными информационными системами для поиска и обработки научно-технической информации и обобщения и ведения научной дискуссии по проблемным вопросам программирования, проектирования и автоматизации.

В процессе работы проводился сравнительный анализ проблем проектирование систем машинного перевода и решений проблемы машинного перевода нейронных сетей в Google Translate.

В результате исследования было выделено решение проблемы машинного перевода нейронных сетей в Google Translate.

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 6](#_Toc26481749)

[1 Выбор направления исследования 7](#_Toc26481750)

[1.1Смысловая проблема 7](#_Toc26481751)

[1.2 Проблема быстрого развития языка 7](#_Toc26481752)

[1.3 Проблема кратковременной памяти 7](#_Toc26481753)

[1.4 Проблема достоверности 9](#_Toc26481754)

[1.5 Сравнительный анализ проблем 10](#_Toc26481761)

[1.6 Вывод 10](#_Toc26481762)

[2 Выбор решения проблемы 11](#_Toc26481763)

[2.2 Анализ предложения 11](#_Toc26481765)

[2.3 Краткий пересказ 11](#_Toc26481767)

[2.4 Присваивание меток 12](#_Toc26481768)

[2.5 Сравнительный анализ решений 12](#_Toc26481769)

[2.6 Вывод 14](#_Toc26481770)

[3 Оценка результатов исследований 15](#_Toc26481771)

[ЗАКЛЮЧЕНИЕ 17](#_Toc26481772)

[СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 18](#_Toc26481773)

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящем отчёте о НИР применяют следующие термины с соответствующими определениями:

Искусственная нейронная сеть (ИНС) - [математическая модель](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C), а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей [нервных клеток](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD) живого организма.

Нейронный машинный перевод Google (GNMT) - система [нейронного машинного перевода](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4) (NMT), разработанная компанией Google и представленная в ноябре 2016 года, которая использует искусственную нейронную сеть для повышения беглости и точности перевода в Google Переводчике

Персональный компьютер - компьютер, предназначенный для эксплуатации одним пользователем, то есть для личного использования.

Машинный перевод с помощью правил (RBMT) - это системы машинного перевода , основанные на лингвистической информации об исходных и целевых языках, в основном получаемой из (одноязычных, двуязычных или многоязычных) словарей и грамматик, охватывающих основные семантические, морфологические и синтаксические закономерности каждого языка соответственно.

Статистический машинный перевод (SMT) - это разновидность машинного перевода текста, основанная на сравнении больших объемов языковых пар.

Гибридный машинный перевод (HMT) - это метод машинного перевода, который характеризуется использованием нескольких подходов машинного перевода в одной системе машинного перевода.

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящем отчете о НИР применяют следующие сокращения и обозначения:

НИР - Научно-исследовательская работа

Нейросеть - Нейронная сеть

ПК - Персональный компьютер

ВВЕДЕНИЕ

Google Translate хоть и является самым популярным и используемым переводчиком в мире, не смог достичь качества перевода профессионального переводчика, даже используя свою новую технологию GNMT.

Но для того, чтобы понять почему нейронные сети не могут решить проблему машинного перевода мы должны рассмотреть какие проблемы у машинного перевода присутствуют, как их решает нейронная сеть и какие изменения мы можем внести в этот процесс дабы улучшить качество перевода или сократить нагрузку на сервера Google.

1. Выбор направления исследования

Ставится задача анализа проблематики выбранной темы НИР по машинному переводу на базе нейронных сетей в Google Translate и проведения сравнительного анализа для выбора самой значимой проблемы.



# **Сравнительный анализ проблем**

При выборе проблемы для рассмотрения необходимо учитывать её выполнимость в данный момент развития технологий. Первая, смысловая проблема является значимой, ключевой проблемой в данный момент для развития нейронных переводческих нейросетей, однако пока что невыполнимой. Далее эта проблема в данной работе рассматриваться не будет.

Также необходимо учитывать затраты ресурсов. При решении проблемы быстрого развития языка, будет тратиться слишком много памяти. Данные растраты не являются благоразумными.

Эту проблему также нельзя считать актуальной, ведь для каждого нового слова можно найти аналог, уже имеющийся в словаре переводчика. Следовательно, эта проблема в данной работе рассматриваться не будет.

Проблема достоверности тоже не является актуальной. Она возникает только либо при бессмысленном, либо при специальном вводе такого количество символов. В этих двух случаях не предполагается нужда в качественном переводе, следовательно, данная проблема в этой работе рассматриваться не будет.

При рассмотрении проблемы кратковременной памяти было выявлено, что эта проблема является актуальной, ведь главная задача переводчика – качественный связный текст, который, как было выяснено в пункте 3, не всегда получается.

Также эта проблема является более выполнимой, чем смысловая проблема, ведь задача состоит не в том, чтобы полностью переучить нейронные сети на литературный перевод, а лишь связать смысл предложений между собой.

# **Вывод**

После проведения сравнительного анализа всех выше рассмотренных проблем, был сделан вывод, что наиболее значимой и выполнимой проблемой в данной области исследования является третья проблема, т.е. проблема кратковременной памяти.

1. Выбор решения проблемы

При анализе данной проблемы было выявлено, что главная задача при её решении – запомнить смысл предыдущего словесного отрезка и учитывать этот смысл при переводе следующего предложения, а также обучить этому нейросеть. На основе этого были выделены следующие примеры решений:



# **Анализ предложения**

Необходимо обучить нейросеть анализу предыдущего предложения и выделению наиболее важных смысловых частей (например, местоимений). Т.е. нейросеть при анализе предложения будет выделять особо важные вещи, такие как окончания глаголов, местоимения, имена. Далее, после анализа и уже зная основной смысл предыдущего предложения (например, что речь идёт о женщине), переводить текущее предложение, используя полученную информацию. В данном примере, необходимо перевести все последующие местоимения в женском роде, а также использовать феминитивы.

# Пример: язык: Малай. Ввод: Dia seorang programmer wanita. Dia bekerja sebagai seorang pengaturcara. Анализ: словосочетание “programmer wanita” переводится как программистка. Следовательно, далее необходимо использовать феминитивы.

# **Краткий пересказ**

Понимать под каждым предложением отдельный текст и запоминать общий смысл, так называемый «краткий пересказ». Использовать данный запомненный краткий пересказ при переводе следующего предложения и далее понимать под текстом уже эти два корректно переведённых предложения. Т.е. необходимо сначала перевести первое предложение, запомнить смысл и составить краткий пересказ (к примеру, речь идёт о женщине Сьюзи, она программист). Далее необходимо проанализировать второе предложение и понять, идёт ли всё ещё речь о том же человеке или объекте, или же начался новый смысловой отрезок. После этого перевести предложение, основываясь на полученной информации предыдущего предложения и текущего. Далее добавить в текст текущее предложение и под текстом понимать уже эти два предложения. Следом необходимо переводить следующее предложение, основываясь на его смысле и на смысле предыдущего текста. Необходимо проводить аналогичные действия со всеми предложениями, пока входящий текст не будет полностью переведён.

Так, каждый раз добавляя в исходных текст новое корректно переведённое предложение, смысл всего исходного текста не будет потерян.

Пример: язык: Малай. Ввод: Dia seorang programmer wanita. Dia bekerja sebagai seorang pengaturcara. John adalah rakan sekerja beliau. Анализ: первое предложение: «она работает программистом», следовательно, при необходимости можно использовать эту информацию в следующем предложении. Второе предложение: т.к. в данном предложении не появляется какой-то новый смысловой отрезок, либо имя, следовательно, рассказ идёт далее про женщину из первого предложения. Значит, «она программистка». Третье предложение: появилось имя «Джон». Следовательно, данное предложение может быть о нём. При переводе и анализе выходит предложение: «Джон – её коллега».

На выходе получится текст: «Она работает программистом. Она программистка. Джон – её коллега», который является корректным, а также где не потерян смысл при переводе местоимений.

# **Присваивание меток**

При переводе предложения необходимо присваивать каждому слову свою специальную метку. Далее, при переводе следующего предложения, необходимо учитывать метки предыдущего и на основе этого анализа делать перевод. Т.е., к примеру, сделать разные метки для «мужчина» и «женщина». Каждый раз, когда нейросеть при переводе будет сталкиваться с именем, должен будет производиться анализ (мужское оно или женское), и на его основе переводится текущее предложение и, если необходимо, следующее. Таким образом, если в первом предложении было использовано имя «Сьюзи», а во втором – слово «sebagai» (которое на малайском языке может означать и «он», и «она»), нейросеть использует метку, выявленную из первого предложения, и переведёт второе с местоимением «она».

# **Сравнительный анализ решений**

* + 1. анализ предложения - было выявлено, что смысл текста всё равно может теряться. Например, если в первом предложении говорится об одной девочке, а во втором – о мальчике. Т.к. нейросеть запомнит информацию о девочке, следующее предложение будет переведено с феминитивами, последствием чего является потеря смысла всего текста.

Также, в данном варианте решения проблемы будет запоминаться основной смысл лишь одного, предыдущего предложения. При таком способе решения связь будет запоминать по «цепочке», что тоже может вызвать проблемы при адекватном и корректном переводе.

При этом, если не проводить такой анализ над текущим предложением, а переводить его с использованием текущий нейросетей, весь смысл будет теряться. Следовательно, данный способ не является эффективным и не решает поставленную задачу корректного и адекватного перевода без потери смысла.

* + 1. краткий пересказ - было выявлено, что запоминание всего предыдущего смыслового отрезка, и, при этом, анализ текущего предложения – самое оптимальное решение. При запоминании смысла всех предыдущих предложений исключается возможность некорректного перевода, а запись этого смысла в краткий пересказ даёт возможность затраты не такого большого количества действующей памяти.
    2. присваивание меток - было выявлено, что при данном решении смысл текста всё равно может теряться. При анализе предложения, нейросеть может ошибиться при расставлении меток, что приведёт к совершенно неправильному переводу.

Также, в данном методе решения не будет анализироваться текущее предложение. В предыдущем и текущем предложениях речь может идти о разных вещах, но, так как нейросеть запомнила метки из предыдущего и не провела анализ текущего, несвязные предложения будут переводится как связные, что приведёт к некорректному переводу.

При сравнительном анализе этих двух вариантов решений было выявлено, что второе решение является самым оптимальным и возможным. С данным решением проблемы кратковременной памяти исходных текст не потеряет свой смысл, а также затраты на память относительно оптимизированы, ведь запоминаться будет не весь предыдущий смысловой отрезок, а лишь его краткий пересказ.

# **Вывод**

Выбрано решение 2 «Краткий пересказ», как самое оптимальное и возможное, с учётом необходимости корректного перевода, а также с учётом минимизации ресурсных затрат.

3 Оценка результатов исследований

При оценке выбранного решения нельзя забывать о практическом тестировании, которое не было проведено в рамках НИР. Основываясь на этом, могут быть предложены следующие этапы внедрения решения с дополнительными рекомендациями по возможным проблемам:

1. предварительная подготовка

Необходимо обеспечить нейросети дополнительной памятью, которая будет использоваться для записи и запоминания кратких пересказов смысловых отрезков. Для этого необходимо примерно проанализировать возможный краткий пересказ при максимально возможном количестве символов и использовать примерно такое количество памяти + некоторое количество для перестраховки, учитывая погрешность.

1. обучение нейросети

Далее необходимо обучить нейросеть новой технологии перевода. Для этого нужно использовать огромное количество текстов (их брать можно, к примеру, с новостных сайтов) и, желательно, на разных языках.

1. закрытое тестирование

Перед тем, как выпускать продукт в массы, необходимо провести закрытое тестирование с профессиональными тестировщиками и переводчиками-людьми. Необходимо сравнивать тексты, переведённые людьми с текстами, переведёнными нейросетью. Далее, при необходимости исправить ошибки.

1. открытое тестирование

Далее можно выпустить продукт на открытое тестирование обычным пользователям. Нейросеть при этом может обучаться от текстов обычных пользователей, а разработчики могут получать обратную связь от пользователей и совершенствовать продукт. Пользователи при этом могут сами решать, какую версию переводчика они хотят использовать.

1. выпуск продукта

Спустя какое время после открытого тестирования можно полностью заменять переводчик со старой нейросетью на переводчик с новой, усовершенствованной нейросетью. При этом необходимо поддерживать продукт, анализируя новый поток данных от пользователей и стараться исправлять ошибки, которые могут появиться.

ЗАКЛЮЧЕНИЕ

Итогом НИР является проанализированная проблема машинного перевода на базе нейронных сетей в Google Translate и найденное решение по улучшению перевода без потери смысла переводчика Google Translate.

Решение содержит этапы и рекомендации по внедрению новых обученных нейронных сетей для более эффективного использования в реальных условиях. При правильном внедрение данного решения в современный Google Translate, должен будет снизиться уровень некорректности перевода из-за текущего уровня обученности нейронных сетей и улучшиться смысловая составляющая переведённых на нужный язык текстов.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

# Машинный перевод: от холодной войны до глубокого обучения [Электронный ресурс]. – Режим доступа: https://vc.ru/future/32616-mashinnyy-perevod-ot-holodnoy-voyny-do-glubokogo-obucheniya (09.11.2019).

# Нейронный машинный перевод Google [Электронный ресурс]. – Режим доступа: https://habr.com/ru/post/414343/ (09.11.2019).

# Wikipedia – the free encyclopedia. Google Translate [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Google\_Translate (09.11.2019).

# Wikipedia – the free encyclopedia. Google Neural Machine Translation [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Google\_Neural\_Machine\_Translation (09.11.2019).

# Что не так с машинным переводом? [Электронный ресурс]. – Режим доступа: https://sysblok.ru/nlp/chto-ne-tak-s-mashinnym-perevodom/ (09.11.2019).

1. Google Translate Doesn’t Work? [Электронный ресурс]. – Режим доступа: https://blog.thelinguist.com/google-translate-doesnt-work (09.11.2019).
2. Wikipedia – the free encyclopedia. Machine translation. [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Machine\_translation (20.11.2019).

# Has AI surpassed humans at translation? Not even close! [Электронный ресурс]. – Режим доступа: https://www.skynettoday.com/editorials/state\_of\_nmt (20.11.2019).

# Небольшие сравнения online-переводчиков [Электронный ресурс]. – Режим доступа: https://habr.com/ru/post/276449/ (20.11.2019).

# Недалёкость Google Translate [Электронный ресурс]. – Режим доступа: https://habr.com/ru/post/410185/ (20.11.2019).

1. Почему переводчикам не нужно бояться нейросетей Гугла [Электронный ресурс]. – Режим доступа: https://habr.com/ru/post/373239/ (20.11.2019).

# Машинное обучение - это весело! Часть 5: Перевод языков с помощью глубокого обучения [Электронный ресурс]. – Режим доступа: https://medium.com/@ppleskov/%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5-%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5-%D1%8D%D1%82%D0%BE-%D0%B2%D0%B5%D1%81%D0%B5%D0%BB%D0%BE-%D1%87%D0%B0%D1%81%D1%82%D1%8C-5-51f40fbae4b0 (20.11.2019).

1. Википедия - свободная энциклопедия. Искусственная нейронная сеть. [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D0%B0%D1%8F\_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F\_%D1%81%D0%B5%D1%82%D1%8C (20.11.2019).
2. Как машинный перевод оценивает… Машина? [Электронный ресурс]. – Режим доступа: https://vk.com/@sysblok-kak-mashinnyi-perevod-ocenivaet-mashina (20.11.2019).
3. Машинный перевод: как это работает? [Электронный ресурс]. – Режим доступа: https://sysblok.ru/nlp/mashinnyj-perevod-kak-jeto-rabotaet/ (20.11.2019).