



Министерство науки и высшего образования Российской Федерации
Калужский филиал
федерального государственного бюджетного
образовательного учреждения высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(КФ МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИУК «Информатика и управление»

КАФЕДРА ИУК4 «Программное обеспечение ЭВМ, информационные технологии»

О Т Ч Е Т

ПРОИЗВОДСТВЕННАЯ ПРАКТИКА

«Технологическая практика»

Студент гр. ИУК4-11М _____ (подпись) (Сафронов Н.С.)
(Ф.И.О.)

Руководитель _____ (подпись) (Белов Ю.С.)
(Ф.И.О.)

Оценка руководителя _____ баллов _____
30-50 (дата)

Оценка защиты _____ баллов _____
30-50 (дата)

Оценка практики _____ баллов _____
(оценка по пятибалльной шкале)

Комиссия: _____ (Гагарин Ю.Е.)
(подпись) (Ф.И.О.)

_____ (Белов Ю.С.)
(подпись) (Ф.И.О.)

_____ (Корнюшин Ю.П.)
(подпись) (Ф.И.О.)

Калуга, 2024

УТВЕРЖДАЮ
Заведующий кафедрой ИУК4
(Гагарин Ю.Е.)
« 05 » сентября 2024 г.

ЗАДАНИЕ **на ПРОИЗВОДСТВЕННУЮ, ТЕХНОЛОГИЧЕСКУЮ ПРАКТИКУ**

За время прохождения практики студенту необходимо:

1. Ознакомиться с существующими электронно-библиотечными системами в сети Интернет, со структурой и содержанием каталогов, протестировать электронные каталоги библиотек, составить перечень литературы по заданной теме, найти полнотекстовые источники, оценить качество электронно-библиотечных ресурсов с позиции пользователя.
2. Ознакомиться с существующими издательскими системами в сети Интернет, получить информацию о возможности свободного доступа к информационным ресурсам российских и зарубежных научных издательств.
3. ознакомиться с порядком составления методических рекомендаций для выполнения работ, выполнить анализ выполняемой работы, разработать методические рекомендации, регламентирующие выполнение работы.
4. получить математическую модель, выполнить анализ полученной модели, выявить факторы, оказывающие существенное влияние на исследуемый процесс, и факторы, позволяющие управлять процессом.
5. выполнить анализ исследуемого технологического процесса, выявить факторы, определяющие характер процесса, разработать план факторного эксперимента, выполнить эксперимент, обработать результаты эксперимента и получить экспериментально-статистическую модель
6. Подготовить отчет и защитить результаты практики.

Дата выдачи задания « 05 » сентября 2024 г.

Руководитель практики _____ Белов Ю.С.

Задание получил
Студент гр. ИУК4-11М _____ Сафронов Н.С.

Оглавление

Введение.....	4
Работа с электронно-библиотечными ресурсами сети Интернет.....	6
Поиск полных текстов научных документов в российских и мировых издательских системах.....	14
Изучение требований и разработка методической документации, регламентирующей выполнение определенного вида работ	20
Построение расчетно-аналитических моделей исследуемых процессов и явлений	23
Планирование и обработка результатов факторных экспериментов	28
Заключение	33
Список литературы	34

Введение

Цели и задачи практики определяются соответствующим самостоятельно устанавливаемым образовательным стандартом МГТУ им. Н.Э. Баумана и рабочей программой практики.

Целью прохождения практики является вклад в формирование компетенций магистранта:

Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий с использованием междисциплинарного подхода, формулировать выводы, адекватные полученным результатам, проводить прогнозирование, ставить исследовательские задачи и выбирать пути их достижения (УКС-1);

Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте (ОПКС-1);

Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями (ОПКС-3);

Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных комплексов (ОПКС-5);

Способен самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности (ОПКС-6);

Способен обоснованно выбирать направления собственных научно-исследовательских и опытно-конструкторских работ, определять цели и задачи научно-практической деятельности в соответствующей профессиональной области (ПКС-7).

Задачи: ознакомиться с существующими электронно-библиотечными системами в сети Интернет, со структурой и содержанием каталогов, протестировать электронные каталоги библиотек, составить перечень литературы по заданной теме, найти полнотекстовые источники, оценить качество электронно-библиотечных ресурсов с позиции пользователя; ознакомиться с существующими издательскими системами в сети Интернет, получить информацию о возможности свободного доступа к информационным ресурсам российских и зарубежных научных издательств, подготовить обзор и создать презентацию информационного характера о периодических изданиях, имеющих в свободном доступе полные тексты статей по заданной теме; ознакомиться с порядком составления методических рекомендаций для выполнения работ, выполнить анализ выполняемой работы, разработать методические рекомендации, регламентирующие выполнение работы; получить математическую модель, выполнить анализ полученной модели, выявить факторы, оказывающие существенное влияние на исследуемый процесс, и факторы, позволяющие управлять процессом; выполнить анализ исследуемого технологического процесса, выявить факторы, определяющие характер процесса, разработать план факторного эксперимента, выполнить эксперимент, обработать результаты эксперимента и получить экспериментально-статистическую модель, провести регрессионный и дисперсионный анализ полученной модели.

Работа с электронно-библиотечными ресурсами сети Интернет

Электронные библиотечные системы (ЭБС) играют ключевую роль в современном образовательном процессе, предоставляя доступ к огромным массивам информации. Они обеспечивают широкий спектр материалов: от научных статей и монографий до учебников и пособий, создавая условия для качественного обучения и научных исследований. ЭБС особенно актуальны в условиях цифровизации образования, когда возникает необходимость в доступе к учебным и научным ресурсам в любое время и из любой точки. В этом обзоре рассмотрены основные ЭБС, используемые в российских образовательных учреждениях, включая Научную электронную библиотеку eLIBRARY.RU, системы «Лань», «Университетская библиотека ONLINE», «IPRbooks» и «Юрайт».

Научная Электронная Библиотека (НЭБ) - E-Library

Научная электронная библиотека eLIBRARY.RU является крупнейшим российским информационным порталом, предоставляющим доступ к рефератам и полным текстам научных статей и публикаций. Библиотека включает более 14 миллионов единиц контента и охватывает все основные научные направления, что делает её важным инструментом для студентов, преподавателей и исследователей.

E-Library предлагает широкий спектр функций, обеспечивающих удобство работы с научной информацией:

- Поиск и фильтрация материалов. Пользователи могут находить публикации по ключевым словам, авторам, названиям журналов и другим параметрам. Это позволяет быстро находить нужные ресурсы среди огромного объёма данных.
- Работа с Российским индексом научного цитирования (РИНЦ). Система предоставляет данные о публикационной активности авторов и

цитировании научных работ, что является важным инструментом для оценки научной значимости и влияния исследований.

- Интеграция с другими системами. E-Library поддерживает экспорт библиографических данных в программы управления источниками, такие как EndNote или Zotero, что упрощает создание ссылок и списков литературы.
- Доступ к полным текстам. Большая часть материалов доступна для скачивания или чтения онлайн, что значительно упрощает процесс подготовки к исследованиям и обучению.

Содержание E-Library охватывает широкий спектр тем, включая технические, естественные, гуманитарные и социальные науки. Эта универсальность делает библиотеку полезной для пользователей с различными образовательными и научными интересами.

Платформа ориентирована на широкую аудиторию. Зарегистрированные пользователи могут:

- Сохранять статьи в избранное для быстрого доступа;
- Настраивать уведомления о новых публикациях в выбранных областях знаний;
- Получать персонализированные рекомендации на основе их поисковой активности.

Также E-Library предлагает институциональные подписки, благодаря которым университеты и научные организации получают доступ к платному контенту.

E-Library активно способствует развитию российской науки, предоставляя доступ к актуальным материалам и инструментам для анализа. Платформа является не только хранилищем научных данных, но и аналитическим инструментом, позволяющим отслеживать динамику научных исследований и строить прогнозы.

Одним из важных аспектов является популяризация российской науки на международном уровне. E-Library поддерживает публикации на

английском языке и индексируется крупными международными базами данных, что повышает видимость работ российских учёных.

Электронно-библиотечная система (ЭБС) издательства "Лань"

Электронно-библиотечная система (ЭБС) издательства "Лань" (Lanbook.ru) является одним из ведущих российских ресурсов, предоставляющих доступ к образовательной и научной литературе. Данная платформа предназначена для студентов, преподавателей, научных сотрудников и специалистов, работающих в различных областях знаний. Система включает электронные версии учебных пособий, монографий, научных журналов и других публикаций, что делает её важным инструментом в образовательном процессе.

ЭБС "Лань" предоставляет пользователям широкий спектр функций, обеспечивающих доступ к актуальной информации:

- Широкий ассортимент литературы. Платформа включает книги и журналы по техническим, естественным, гуманитарным и другим наукам. Это делает её универсальным источником знаний для различных категорий пользователей.
- Интерактивный поиск. Удобная система поиска позволяет находить материалы по названию, автору, ключевым словам, а также по разделам и дисциплинам.
- Персонализированные настройки. Пользователи могут создавать личные кабинеты, сохранять избранные материалы и настраивать уведомления о выходе новых изданий в интересующей их области.
- Доступ к образовательным курсам. На платформе представлены материалы, интегрированные в учебные программы, что упрощает подготовку к занятиям и экзаменам.
- Интеграция с LMS. ЭБС "Лань" поддерживает взаимодействие с системами управления обучением (Learning Management Systems), что

облегчает преподавателям и студентам процесс доступа к учебным материалам.

Каталог ЭБС "Лань" структурирован таким образом, чтобы обеспечить удобную навигацию по различным тематическим разделам. Большая часть контента представлена в виде учебников и учебных пособий, что делает ресурс особенно полезным для студентов. Кроме того, система включает монографии и научные журналы, предназначенные для исследовательской работы.

ЭБС "Лань" предоставляет как индивидуальные, так и институциональные подписки. Благодаря этому студенты и преподаватели учебных заведений могут получить доступ к ресурсам, необходимым для обучения и научной работы.

ЭБС "Лань" играет ключевую роль в образовательной системе России, предоставляя доступ к качественным источникам знаний. Этот ресурс активно используется в университетах и колледжах благодаря интеграции с учебными планами. Система помогает образовательным учреждениям внедрять цифровые технологии, обеспечивая доступ к актуальным материалам.

Кроме того, платформа способствует развитию научной деятельности. Публикации монографий и научных статей на платформе повышают их доступность для исследователей, тем самым способствуя обмену знаниями и междисциплинарному сотрудничеству.

Электронно-библиотечная система (ЭБС) «Университетская библиотека ONLINE»

Электронно-библиотечная система (ЭБС) «Университетская библиотека ONLINE» (biblioclub.ru) представляет собой универсальную цифровую платформу, предоставляющую доступ к разнообразным образовательным и научным ресурсам. Данный ресурс ориентирован на удовлетворение потребностей студентов, преподавателей, научных работников и библиотек, предлагая обширный каталог материалов для учебной, научной и профессиональной деятельности. ЭБС является одной из наиболее

востребованных систем в образовательной среде благодаря широкому охвату тематики и постоянному обновлению фонда.

ЭБС «Университетская библиотека ONLINE» выделяется рядом функциональных возможностей, которые делают её удобной и полезной для пользователей:

- Широкий тематический охват. Система предоставляет доступ к учебникам, учебным пособиям, монографиям, научной периодике и другим изданиям по всем основным направлениям подготовки: от естественных и технических наук до гуманитарных дисциплин.
- Интерактивные возможности. Пользователи могут не только читать материалы, но и выделять текст, делать заметки, закладки и сохранять избранное для дальнейшего использования.
- Доступ в любое время и из любой точки. Онлайн-доступ к материалам обеспечивает удобство работы независимо от местоположения пользователя.
- Интеграция с образовательным процессом. В системе представлены материалы, рекомендованные учебными программами, что делает её важным элементом в подготовке к занятиям и выполнению исследовательских проектов.
- Адаптация для разных категорий пользователей. Интерфейс системы интуитивно понятен и адаптирован для использования как студентами, так и преподавателями.

Особое внимание уделено качеству контента, что достигается за счёт сотрудничества с ведущими российскими издательствами и авторами. Каталог регулярно обновляется, что обеспечивает его актуальность и соответствие современным требованиям.

ЭБС «Университетская библиотека ONLINE» активно используется в образовательных учреждениях России для поддержки учебного процесса. Университеты и колледжи подключают своих студентов к системе, что позволяет им иметь доступ к необходимым учебникам и пособиям в

электронном виде. Это особенно важно в условиях перехода на цифровые технологии обучения, включая дистанционное и смешанное обучение.

Электронно-библиотечная система (ЭБС) «IPRbooks»

Электронно-библиотечная система (ЭБС) «IPRbooks» (iprbookshop.ru) является одной из ведущих образовательных платформ в России, предоставляющих доступ к информационным ресурсам для учебного и научного процесса. Основной задачей системы является обеспечение студентов, преподавателей и исследователей качественными материалами для всех этапов обучения, от бакалавриата до аспирантуры, а также для повышения квалификации.

ЭБС «IPRbooks» сочетает в себе широкий спектр возможностей, которые делают её эффективным инструментом для образовательного процесса:

- **Разнообразие контента.** В базе данных содержатся учебники, монографии, научные статьи, журналы и другие материалы, охватывающие более 10 000 наименований. Эти ресурсы соответствуют требованиям государственных образовательных стандартов.
- **Универсальный доступ.** Система позволяет работать с материалами из любой точки мира благодаря онлайн-доступу. Кроме того, пользователи могут загружать книги для работы в офлайн-режиме через специальное приложение.
- **Адаптация к образовательным потребностям.** Материалы в «IPRbooks» сгруппированы по направлениям подготовки, что упрощает поиск литературы, соответствующей учебной программе.
- **Многоязычность.** В дополнение к русскоязычным ресурсам платформа включает материалы на иностранных языках, что делает её полезной для изучения международных дисциплин.

Одной из уникальных особенностей является сотрудничество с более чем 200 ведущими российскими издательствами, которые регулярно пополняют фонд актуальными изданиями. Это делает платформу динамичной и актуальной для образовательной среды.

ЭБС «IPRbooks» занимает важное место в системе высшего и дополнительного профессионального образования России. Её ресурсы активно используются студентами и преподавателями для выполнения различных задач:

Электронно-библиотечная система (ЭБС) «Юрайт»

ЭБС «Юрайт» обладает рядом уникальных характеристик, которые делают её востребованной в образовательной среде:

- Широкий ассортимент литературы. Коллекция включает более 3000 наименований, основную часть из которых составляют учебники и учебные пособия. Все издания соответствуют актуальным требованиям Федеральных государственных образовательных стандартов (ФГОС).
- Регулярное обновление контента. Фонд системы постоянно пополняется новыми изданиями, что позволяет пользователям получать доступ к самой актуальной информации.
- Доступность для всех уровней образования. В системе представлены материалы для бакалавриата, магистратуры, аспирантуры, а также для повышения квалификации и профессионального обучения.

Основу коллекции «Юрайт» составляют издания одноимённого издательства, известного своим качественным подходом к разработке образовательной литературы. Помимо учебников и учебных пособий, в системе можно найти монографии, практические руководства, сборники задач и тестов, а также дополнительные материалы для самостоятельного обучения.

ЭБС «Юрайт» активно используется как инструмент поддержки учебного процесса. Система также позволяет преподавателям формировать

индивидуальные учебные планы, добавляя в них материалы из библиотеки. Это облегчает процесс подготовки к занятиям и способствует повышению качества обучения.

Для университетов, колледжей и других образовательных организаций ЭБС «Юрайт» предлагает готовое решение для обеспечения студентов и преподавателей учебными материалами.

Поиск полных текстов научных документов в российских и мировых издательских системах

Для достижения целей и задач, поставленных в рамках исследования по теме "Генерация программного кода на основе естественного языка", необходимо было изучить и проанализировать научные документы, касающиеся ключевых аспектов данной области. Особое внимание уделялось работам, описывающим следующие направления:

- Основы применения моделей глубокого обучения для преобразования естественного языка в программный код.
- Архитектуры трансформеров, используемые в современных системах генерации кода.
- Анализ существующих датасетов, применяемых для обучения моделей генерации кода, включая их состав, особенности и применимость.
- Методы оценки качества сгенерированного кода, такие как BLEU, ROUGE, METEOR и другие.
- Текущее состояние технологий и примеры их использования в реальных проектах, а также перспективы развития данной области.

Для сбора информации и поиска релевантных публикаций использовались как российские, так и мировые электронно-библиотечные системы. Анализируемые материалы включали научные статьи, монографии, диссертации, а также материалы конференций, посвящённых искусственному интеллекту, обработке естественного языка и программной инженерии.

Dehaerne, E., Dey, B., Halder, S., De Gendt, S., & Meert, W. (2022). Code Generation Using Machine Learning: A Systematic Review. IEEE Access.

В статье представлен систематический обзор существующих методов генерации кода с использованием машинного обучения. Основное внимание уделено трем типам задач: генерация кода из естественного языка (description-

to-code), создание документации из кода (code-to-description) и модификация существующего кода (code-to-code). Авторы анализируют различные модели машинного обучения, включая RNN, LSTM, transformers и CNN, а также подходы к токенизации и методы оценки качества синтезированного кода.

Приложения для генерации кода:

- Генерация кода из текстовых описаний.
- Автоматическое создание документации.
- Перевод кода между языками программирования и исправление ошибок.

Модели машинного обучения:

- LSTM и RNN используются для учета длинных зависимостей в данных.
- Transformers эффективно работают с естественным языком и кодом.
- CNN применяются для задач, связанных с изображениями.

Методы токенизации:

Word-based, character-based и subword-based токенизация с акцентом на последний метод, который помогает справляться с проблемами ограниченного словаря.

Оценка синтезированного кода:

- Метрики совпадений токенов, такие как BLEU, CIDEr, ROUGE и METEOR.
- Анализ функциональной правильности с использованием динамического и статического анализа.

Статья представляет собой всесторонний обзор современных подходов к генерации кода с применением машинного обучения и подчеркивает важность выбора правильных моделей и методов токенизации. Работы в этой области продолжают развиваться, и предложенные направления исследований помогут улучшить эффективность и точность генерации кода. Статья будет

полезна для исследователей и разработчиков, занимающихся автоматическим программированием и машинным обучением.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems

В статье "Attention Is All You Need" авторы представляют революционную архитектуру трансформера, которая стала основой для современных моделей обработки естественного языка, включая системы генерации программного кода. Основной инновацией работы является введение механизма самовнимания (self-attention), который заменяет рекуррентные и свёрточные слои, значительно упрощая архитектуру модели и повышая её производительность.

В статье детально рассматриваются ключевые элементы трансформера: слои самовнимания, механизмы нормализации и позиционного кодирования, а также применение этих компонентов в задачах перевода текста. Авторы демонстрируют эффективность предложенной модели на задаче машинного перевода, показывая, что трансформер превосходит существующие методы по скорости обучения и качеству результата.

Работа предназначена для исследователей и разработчиков, занимающихся глубоким обучением и обработкой естественного языка. Она закладывает основу для дальнейших разработок в области трансформерных архитектур и их применения в задачах генерации текста, включая генерацию программного кода.

Chen, M., et al. (2021). Evaluating Large Language Models Trained on Code.

В статье рассматриваются возможности языковой модели Codex, специализированной на генерации программного кода, разработанной на

основе GPT-3 и обученной на коде из открытых репозиториях GitHub. Основное внимание уделено задаче синтеза функций Python по текстовым описаниям (docstrings) и оценке функциональной корректности кода с использованием набора тестов HumanEval.

Модель Codex достигает существенного улучшения в точности синтеза кода по сравнению с другими моделями (например, GPT-3 и GPT-J). Также исследуются такие методы, как многократная генерация выборок для повышения качества решений. Среди ограничений модели выделены сложности с длинными цепочками операций и их привязкой к переменным.

Статья включает обсуждение этических, экономических и экологических последствий применения подобных технологий, а также возможных рисков, связанных с безопасностью и злоупотреблением. Работа представляет интерес для разработчиков, инженеров в области машинного обучения и специалистов по генерации кода.

Gu, A., & Dao, T. (2023). Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In Advances in Neural Information Processing Systems.

В статье "Mamba: Linear-Time Sequence Modeling with Selective State Spaces" авторы Альберт Гу и Три Дао представляют новую архитектуру модели Mamba, которая обеспечивает линейное масштабирование по длине последовательности и высокую производительность на различных типах данных.

Модели на основе трансформеров страдают от квадратичной сложности вычислений на длинных последовательностях. Линейные модели внимания, рекуррентные нейронные сети (RNN) и свёрточные модели не всегда достигают качества трансформеров на важных типах данных, таких как текст.

Mamba успешно решает задачи, требующие контекстно-зависимого рассуждения, такие как "Selective Copying" и "Induction Heads", и может экстраполировать решения на последовательности длиной более 1 миллиона

токенов. Mamba превосходит трансформеры по качеству предсказаний и скорости генерации текста, показывая лучшие результаты на задачах обработки естественного языка, включая общие тесты на здравый смысл.

Представленная архитектура Mamba является первой моделью, которая линейно масштабируется по длине последовательности и достигает качества трансформеров на различных типах данных.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. In Advances in Neural Information Processing Systems (NIPS).

В статье представлена новая архитектура нейронных сетей для решения задач, требующих работы с последовательностями данных, таких как машинный перевод. Авторы предлагают использование многослойной архитектуры LSTM для кодирования и декодирования последовательностей. Ключевыми инновациями являются использование LSTM для преобразования входной последовательности в фиксированное представление и применение реверсирования порядка слов для улучшения результатов.

Метод LSTM значительно улучшает качество машинного перевода на наборе данных WMT'14, достигая BLEU-оценки 34.81, что превосходит традиционные статистические методы.

Пересортировка гипотез с использованием LSTM повышает точность перевода до 36.5.

Применение реверсирования порядка слов помогает справляться с долгими зависимостями в длинных предложениях.

Работа открывает новые возможности для решения сложных задач с переменной длиной входных и выходных последовательностей и является важным вкладом в области обработки естественного языка.

Ghosh, D., Zhu, P. H., & Bayley, I. (2024). Benchmarks and Metrics for Evaluations of Code Generation: A Critical Review. arXiv preprint arXiv:2400.08768.

В статье проведен критический обзор существующих подходов к оценке моделей генерации программного кода, с фокусом на метрики и бенчмарки. Авторы анализируют различные типы задач программирования, такие как описание в код (D2C), код в описание (C2D) и код в код (C2C), а также модели больших языковых моделей (LLMs) для программирования, такие как GPT-NEO, GPT-J и Codex. Рассматриваются проблемы с текущими метриками качества, такими как синтаксическая близость (BLEU, ROUGE-L) и функциональная корректность. Также подчеркивается важность создания разнообразных и объективных бенчмарков для оценки моделей, а также разработки новых метрик, которые учитывают удобство использования и продуктивность моделей. В статье предлагаются направления для будущих исследований, включая улучшение тестовых наборов и использование метаданных для сценарных оценок.

Изучение требований и разработка методической документации, регламентирующей выполнение определенного вида работ

Основные требования по оформлению учебных пособий

При подготовке методической документации следует соблюдать требования, указанные в приложениях к ГОСТам, таким как ГОСТ 2.105-95 ЕСКД, ГОСТ 2.106-96 ЕСКД, ГОСТ 7.32-2001, ГОСТ 7.9-95, ГОСТ 7.12-1993, ГОСТ 7.54-88, ГОСТ 7.1-2003 и ГОСТ Р 7.0.5-2008. Эти стандарты регулируют основные требования к текстовым документам, структуре и оформлению отчетов о научных исследованиях, а также к правилам оформления рефератов и аннотаций, сокращений на русском языке, представлению числовых данных, библиографической записи и ссылок.

Методическая документация должна быть составлена и оформлена в соответствии с вышеупомянутыми ГОСТами и установленными нормами. Для текста рекомендуется использовать шрифт Times New Roman размером 14 пунктов на бумаге формата А4, обеспечивая четкость и аккуратность изложения. Следует использовать автоматическую расстановку переносов, соблюдать требования к оформлению текста, заголовков, таблиц и рисунков. Важно также избегать внедрения данных из других документов MS Word и соблюдать грамотность и доступность изложения.

При оформлении документа необходимо учитывать правильные размеры полей, выравнивание текста, отступы, интервалы и расстояния между заголовками и основным текстом, а также последовательность нумерации страниц. Титульный лист является первой страницей, но номер страницы на нем не ставится, хотя он включается в общую нумерацию. Бланки титульных листов для различных типов выпускных квалификационных работ можно найти в приложениях к положениям о подготовке и защите ВКР.

Нумерация страниц начинается с титульного листа, но сам титульный лист не нумеруется. Нумерация страниц должна располагаться сверху или внизу

страницы по центру. Заголовки не нумеруются и выравниваются по центру страницы. Все иллюстрации должны иметь ссылки в тексте, быть выровнены по центру и иметь номер и описание. Пояснительный текст к иллюстрациям располагается под ними.

Оценка знаний, умений и навыков

Учебные, методические и учебно-методические материалы, прошедшие редакционно-издательскую обработку и получившие статус издания, используются для распространения информации и поддержания учебного процесса. Эти материалы бывают в виде печатных и электронных изданий. Учебное издание представляет собой систематизированные научные или прикладные сведения, изложенные в доступной для изучения форме, и предназначено для учащихся различных возрастных групп.

Образовательная программа описывает основные характеристики образования, организационно-педагогические условия, формы аттестации и другие элементы, необходимые для реализации программы.

Учебный план включает перечень дисциплин, аттестационных испытаний и других видов учебной деятельности, с указанием их объема и последовательности. Для каждой дисциплины и практики указываются формы промежуточной аттестации студентов.

Учебно-методический комплекс документации (УМКД) представляет собой систему нормативных и учебно-методических документов, которые определяют цели, содержание и методы обучения студентов в рамках конкретной дисциплины. Включает программу дисциплины, курс лекций, методические указания, задания для самостоятельной работы, контрольно-оценочные средства и другие материалы.

Программа учебной дисциплины — это документ, который включает перечень ожидаемых результатов обучения, указания на место дисциплины в образовательной программе, объем дисциплины, содержание, разделенное на темы, перечень материалов для самостоятельной работы, фонд оценочных

средств для промежуточной аттестации, а также основную и дополнительную литературу и методические указания.

Программа практики — это документ, который описывает вид, метод и форму проведения практики, перечень ожидаемых результатов обучения, место практики в образовательной программе, объем, содержание, формы отчетности, оценочные средства, литературу, ресурсы Интернета, информационные технологии и материально-техническую базу.

Программа государственной итоговой аттестации — это документ, который включает вопросы для государственного экзамена, требования к выпускным квалификационным работам и рекомендованную литературу для подготовки.

Учебник — это учебное издание, которое содержит систематизированное изложение дисциплины, соответствует учебной программе и является основным изданием по дисциплине. Он охватывает базовые знания, обязательные для освоения студентами, и соответствует образовательному стандарту. Учебник предназначен для самостоятельного освоения его содержания.

Учебное пособие — это учебное издание, которое дополняет или заменяет учебник, официально утверждено и содержит не менее 15 печатных листов. К учебным пособиям относятся курс лекций, конспект лекций и рабочая тетрадь.

Учебное издание включает систематизированные научные и практические сведения, изложенные в удобной для изучения форме.

Практикумы — это учебные издания, содержащие практические задания и упражнения для закрепления пройденного материала.

Построение расчетно-аналитических моделей исследуемых процессов и явлений

В процессе научных исследований, а также в приложениях различных областей науки и техники, важным этапом является построение расчетно-аналитических моделей, которые позволяют прогнозировать, анализировать и интерпретировать исследуемые процессы и явления. Моделирование представляет собой методику, основанную на формализации реальных процессов в виде математических или алгоритмических конструкций, что дает возможность исследовать их поведение в различных условиях. Важно отметить, что расчетно-аналитические модели служат связующим звеном между теоретическими выводами и практическими приложениями, обеспечивая точность, обоснованность и возможность повторяемости научных исследований.

Генерация программного кода на основе естественного языка является одной из самых актуальных и перспективных задач в области искусственного интеллекта и машинного обучения. Суть задачи заключается в автоматическом преобразовании текстовых описаний, данных на естественном языке, в работающий программный код. Это направление имеет огромный потенциал для упрощения разработки программного обеспечения, автоматизации рутинных задач и улучшения взаимодействия между людьми и машинами, особенно в контексте людей, не обладающих глубокими знаниями в области программирования.

С развитием искусственного интеллекта и, в частности, моделей глубокого обучения, задача генерации программного кода стала значительно более выполнимой. Модели на основе нейронных сетей, такие как Sequence-to-Sequence (Seq2Seq) и трансформеры, оказались особенно эффективными для решения таких задач. Эти модели обучаются на больших объемах данных, содержащих пары "описание — код", и могут генерировать программный код по заданному текстовому описанию. Применение таких моделей позволяет

улучшить продуктивность разработки, ускорить процесс программирования и повысить доступность программирования для широкой аудитории.

Модель Sequence-to-Sequence для генерации текстовых последовательностей

Модель Sequence-to-Sequence (Seq2Seq) является одной из основополагающих архитектур в области глубокого обучения для задач, в которых входные и выходные последовательности имеют различную длину. Первоначально разработанная для задач машинного перевода, эта модель была адаптирована для множества других приложений, включая задачи генерации программного кода. Архитектура Seq2Seq позволяет эффективно моделировать сложные зависимости в данных, что делает её идеальной для преобразования текстовых описаний задач в соответствующий программный код.

Seq2Seq состоит из двух частей: кодировщика (encoder) и декодировщика (decoder), часто реализуемых с использованием рекуррентных нейронных сетей (RNN), таких как LSTM или GRU. Кодировщик преобразует входную последовательность в фиксированное представление, называемое контекстным вектором, который затем используется декодировщиком для создания выходной последовательности.

Однако стандартная Seq2Seq модель имеет ограничение: использование одного фиксированного контекстного вектора для всех данных входной последовательности не всегда эффективно, особенно для длинных последовательностей.

Для решения проблемы фиксированного контекстного вектора был предложен механизм внимания, который позволяет декодировщику динамически фокусироваться на наиболее важных частях входной последовательности. Вместо того чтобы использовать один контекстный вектор, внимание позволяет вычислять веса для каждого элемента входной последовательности, повышая точность генерации.

В задаче генерации программного кода механизм внимания помогает модели точно обрабатывать зависимости между различными частями программы, улучшая точность и качество сгенерированного кода. Это делает модель Seq2Seq с вниманием эффективным инструментом для трансформации текстовых описаний в программный код.

Модели генеративного предварительно обученного трансформера (GPT) для генерации текстовых последовательностей

Модели генеративного предварительно обученного трансформера (GPT) являются мощным инструментом для генерации текстовых последовательностей. Они используют архитектуру трансформера, которая представляет собой модель, основанную на механизме внимания, и обеспечивают высокую производительность в задачах обработки естественного языка.

Основной компонент модели GPT — это трансформер, который состоит из двух частей: кодировщика и декодировщика. Однако в GPT используется только декодировщик, что отличает его от моделей Seq2Seq. Весь процесс обучения и генерации текста происходит в одном направлении — от начала последовательности к концу.

Механизм внимания в трансформере позволяет модели учитывать все элементы входной последовательности одновременно, используя несколько слоев внимания. Это позволяет эффективно захватывать долгосрочные зависимости между словами, что критично для генерации осмысленных и последовательных текстов.

GPT использует стратегию предварительного обучения на больших объемах текста, где модель обучается предсказывать следующее слово в последовательности, основываясь на предыдущем контексте. Этот процесс позволяет модели усваивать общие языковые паттерны и грамматические структуры.

После предварительного обучения GPT может быть дообучена на специализированных данных для решения конкретных задач, таких как генерация программного кода или других специфичных типов текстов.

Одним из ключевых преимуществ GPT является способность моделировать долгосрочные зависимости, что делает её эффективной для генерации длинных и сложных текстовых последовательностей. Кроме того, её архитектура позволяет обрабатывать текст с высокой степенью контекстуальной осведомленности, что улучшает качество сгенерированного контента.

Модели GPT, обученные на больших корпусах программного кода, могут генерировать код по текстовым описаниям задач или даже продолжать фрагменты кода, обеспечивая высокое качество и точность результатов. Важно, что GPT может работать как с природным языком, так и с синтаксисом программирования, что расширяет её применение в задачах автоматической генерации программного обеспечения.

Примером такого применения является генерация программных решений по текстовым инструкциям или исправление ошибок в коде. GPT использует контекст предыдущих операций и создаёт логику, соответствующую задуманному функционалу программы.

Модели Mamba для обработки длинных последовательностей

Архитектура Mamba является одним из новейших достижений в области глубокого обучения и направлена на решение проблемы обработки длинных последовательностей данных, которая остается актуальной даже с появлением мощных трансформеров. Mamba предлагает уникальный подход, основанный на структурированных моделях состояний (State-Space Models, SSM), который сочетает высокую производительность с эффективным использованием вычислительных ресурсов.

Современные задачи, такие как генерация кода, анализ текста или биологических последовательностей, требуют обработки длинных входных

данных. Трансформеры, являясь стандартом в этой области, сталкиваются с рядом ограничений:

Механизм самовнимания (self-attention) масштабируется квадратично относительно длины входной последовательности, что приводит к значительному увеличению вычислительных затрат при увеличении длины.

Обработка длинных последовательностей требует значительных объемов оперативной памяти для хранения промежуточных представлений.

Несмотря на высокую точность, трансформеры могут быть неэффективными при работе с задачами, требующими учета широкого контекста.

Архитектура Mamba разрабатывалась с целью решения этих проблем, обеспечивая баланс между точностью, эффективностью и скоростью работы.

В архитектуре Mamba используется динамическое кодирование позиций, которое обеспечивает адаптивное представление порядка токенов. В отличие от фиксированных позиционных эмбедингов в трансформерах, динамический подход позволяет более точно учитывать длинные зависимости, особенно в задачах, где порядок элементов имеет решающее значение.

Каждое новое состояние вычисляется на основе предыдущего, что даёт модели возможность учитывать историю последовательности. Это особенно важно для задач, где текущие действия зависят от ранее обработанных данных, например, при автодополнении кода.

Различные подходы к моделированию, включая модели Seq2Seq, GPT и Mamba, позволяют эффективно решать задачи, связанные с обработкой последовательностей данных. Эти модели обеспечивают высокую точность предсказаний, а также оптимизируют вычислительные ресурсы, что особенно важно для работы с большими объемами данных. Применение современных методов моделирования позволяет улучшить качество решений в таких областях, как генерация программного кода, анализ текстов и обработка сложных зависимостей в данных.

Планирование и обработка результатов факторных экспериментов

В последние годы область генерации программного кода из естественного языка привлекает всё большее внимание как со стороны исследователей, так и со стороны разработчиков, стремящихся упростить процесс программирования и сделать его доступным для широкого круга пользователей. Одной из ключевых проблем в данной области является создание эффективных моделей, способных интерпретировать естественный язык и преобразовывать его в корректный и оптимизированный код. Однако для достижения высокой точности и эффективности этих моделей необходимы продуманные и систематические подходы, такие как факторные эксперименты.

Факторные эксперименты представляют собой методику, позволяющую оценивать влияние различных факторов (например, архитектуры модели, гиперпараметров, структуры входных данных) на результаты выполнения моделей. В контексте генерации программного кода из естественного языка эти эксперименты помогают оптимизировать процесс выбора наилучшей модели, а также дают возможность выявить важнейшие факторы, влияющие на точность и эффективность генерации.

Планирование факторных экспериментов включает в себя выбор факторов, которые могут повлиять на процесс генерации кода, и их сочетание в рамках экспериментального исследования. Для эффективного планирования важно определить, какие аспекты модели требуют оптимизации и какие факторы наиболее существенно влияют на качество работы модели.

Одним из ключевых аспектов при планировании эксперимента является выбор факторов, которые могут включать следующие параметры:

- Тип модели (например, Seq2Seq, GPT, Mamba);
- Длина входной последовательности (например, текстовая длина);
- Тип входных данных (например, простые команды или сложные запросы);

- Предобработка данных (токенизация, нормализация текста);
- Гиперпараметры модели (например, размер скрытого состояния, скорость обучения);
- Методы оценки качества генерации кода (например, BLEU, ROUGE, точность синтаксической корректности).

Планирование эксперимента должно учитывать не только эти параметры, но и возможные их взаимодействия. Например, выбор модели может зависеть от длины входной последовательности, что также влияет на эффективность работы модели. Следовательно, важно провести серию экспериментов с учетом этих взаимодействий, чтобы оценить влияние каждого фактора на итоговые результаты.

После определения факторов и их значений необходимо провести серию экспериментов для оценки влияния каждого из них на качество генерации кода. Важным шагом является выбор метода анализа и обработки результатов, который поможет выявить наиболее значимые факторы.

Структура эксперимента включает в себя выбор вида эксперимента, который может быть полным факторным или частичным факторным. Полный факторный эксперимент предполагает исследование всех возможных комбинаций факторов, что может быть вычислительно затратным при большом количестве факторов. В таких случаях целесообразно использовать частичный факторный эксперимент, где только часть всех возможных комбинаций исследуется с целью уменьшения вычислительных затрат.

Методология проведения экспериментов в контексте генерации программного кода из естественного языка может включать следующие шаги:

- Разбиение данных на тренировочную и тестовую выборки для каждого типа модели.
- Обучение моделей на тренировочных данных с использованием различных значений гиперпараметров.
- Генерация программного кода из текстовых запросов на основе обученных моделей.

Оценка качества сгенерированного кода с помощью стандартных метрик качества, таких как точность, полнота, и синтаксическая корректность.

Проведение экспериментов позволяет собрать статистику по каждому из факторов, чтобы определить, какой из них оказывает наибольшее влияние на конечные результаты.

После проведения экспериментов необходимо провести обработку и анализ собранных данных, чтобы оценить влияние каждого фактора на результат. Для этого могут быть использованы различные статистические методы, такие как анализ дисперсии (ANOVA), корреляционный анализ и методы машинного обучения для классификации факторов по важности.

Статистическая обработка результатов предполагает использование методов, которые позволяют выявить значимость каждого фактора. Например, анализ дисперсии позволяет оценить, насколько изменения в каждом факторе влияют на итоговый результат, и определить, какие факторы являются наиболее важными. Для повышения точности анализа можно применить методики, такие как градиентный бустинг или случайные леса, которые могут выявить сложные зависимости между факторами.

Графический анализ результатов также имеет большое значение. Построение графиков зависимости метрик качества генерации кода от различных факторов позволяет визуализировать, какие изменения в факторах ведут к улучшению или ухудшению результатов. Это помогает быстро выявить закономерности и выбрать оптимальные параметры.

Планирование и обработка результатов факторных экспериментов являются важнейшими этапами в процессе разработки эффективных моделей генерации программного кода из естественного языка. Применение методов факторных экспериментов позволяет оптимизировать гиперпараметры моделей, выявить наиболее важные факторы, которые влияют на качество генерации, и выбрать наилучшие методы для реализации задач генерации кода.

Для оценки и сравнения эффективности моделей генерации программного кода используются специализированные бенчмарки и датасеты.

Эти ресурсы предоставляют множество примеров задач, на которых тестируются различные подходы и архитектуры, что позволяет глубже понять их преимущества и ограничения.

Одним из самых популярных наборов данных является CodeXGLUE. Это обширный датасет, включающий данные для множества различных задач, таких как автодополнение кода, перевод с одного языка программирования на другой и анализ ошибок. Он охватывает широкий спектр языков программирования и применяется для тренировки трансформеров и других моделей машинного обучения.

Другим важным датасетом является CoNaLa (Code/Natural Language Challenge), который предназначен для генерации кода на языке Python по запросам на естественном языке. Этот набор данных включает около 20 тысяч пар запросов и соответствующего кода, что позволяет моделям обучаться на реальных примерах и улучшать способность понимать и генерировать программный код по текстовому описанию.

Для решения задач, связанных с анализом и генерацией кода на языке Python, активно используется Py150. Этот датасет состоит из более 150 тысяч примеров, что позволяет моделям развивать навыки предсказания ошибок, автодополнения и рефакторинга кода.

Еще одним полезным ресурсом является CodeSearchNet, датасет, который включает 6 миллионов примеров кода на различных языках программирования, таких как Python, Java, JavaScript и другие. Он предназначен для задач поиска по коду и генерации описаний кода, что делает его важным инструментом для разработки более универсальных моделей.

Эти датасеты являются неотъемлемой частью экосистемы машинного обучения для генерации программного кода. Они предоставляют широкие возможности для тестирования различных моделей и подходов, а также служат основой для дальнейшего совершенствования моделей, обученных на реальных примерах.

Результаты, полученные на основе этих экспериментов, демонстрируют, что правильный выбор модели и гиперпараметров может существенно

повысить точность и эффективность генерации кода. В дальнейшем такие эксперименты помогут не только улучшить существующие подходы, но и разработать новые методы для создания более мощных и гибких систем генерации программного кода из естественного языка.

Заключение

Для грамотного построения работы над темой генерации программного кода с использованием современных подходов машинного обучения необходимо проводить анализ различных материалов, которые следует тщательно изучить. Для этого активно используются электронно-библиотечные системы (ЭБС), предоставляющие доступ к широкому спектру научных публикаций, статей и исследований.

Работа с ЭБС в сети Интернет позволяет быстро и удобно находить необходимые данные, читать и использовать электронные издания как в образовательных, так и в личных целях. Основные преимущества использования ЭБС включают доступность, удобство работы, широкую функциональность, обширную коллекцию материалов и экономическую целесообразность.

В результате анализа открытых источников было выявлено, что применение машинного обучения для генерации программного кода является актуальной и развивающейся областью информационных технологий. Рассмотрены ключевые подходы, такие как seq2seq, GPT, и Mamba, а также методы токенизации и оценки качества синтезированного кода.

Основными преимуществами таких подходов являются повышение автоматизации процессов разработки, улучшение качества кода и ускорение разработки программных продуктов. Однако, остаются определенные проблемы, такие как необходимость улучшения точности генерируемого кода и повышение его функциональной правильности.

В заключение следует отметить, что использование машинного обучения для генерации кода является перспективным и активно развивающимся направлением, способным существенно улучшить процессы разработки программного обеспечения, повысить производительность и упростить создание сложных программных решений.

Список литературы

Литература по практике

1. Моделирование информационных ресурсов [Электронный ресурс]: учебно-методический/ Составитель Огнев Э.Н. - Кемерово : Кемеровский государственный университет культуры и искусств, 2013. - 36 с. : ил., табл. - URL: <http://biblioclub.ru/index.php?page=book&id=274218>
2. Коваленко, Ю.В. Информационно-поисковые системы [Электронный ресурс]: учебно-методическое пособие / Ю.В. Коваленко, Т.А. Сергиенко. — Омск: Омская юридическая академия, 2017. — 38 с.— Режим доступа: <http://www.iprbookshop.ru/66817.html>
3. Рыжков, И.Б. Основы научных исследований и изобретательства[Электронный ресурс] : учебное пособие / И.Б. Рыжков. — 3-е изд., стер. — Санкт-Петербург : Лань, 2019. — 224 с.— URL: <https://e.lanbook.com/book/116011>
4. Антонов, А.В. Системный анализ [Текст]: учебник / А.В. Антонов. М.: Высш. шк., 2006. – 454 с.
5. Рогов, В.А. Методика и практика технических экспериментов[Текст] : учеб. пособие / В.А.Рогов, А.В. Антонов, Г.Г. Поздняк. – М.: Академия, 2005. – 288 с.