**Министерство науки и высшего образования Российской Федерации**

**Федеральное государственное автономное образовательное учреждение высшего образования**

**«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

КАФЕДРА ТЕХНОЛОГИЙ ПРОГРАММИРОВАНИЯ

Направление: 09.03.03 – «Прикладная информатика»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**Разработка алгоритма обработки естественного языка при помощи машинного обучения с использованием нейросетей BERT**

**Работа завершена:**

Студент 4 курса

группы 09-852

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_ 2022 г. \_\_\_\_\_\_\_\_\_\_\_\_ Саидмуродов С.С.

**Работа допущена к защите:**

Научный руководитель

старший преподаватель

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_ 2022 г. \_\_\_\_\_\_\_\_\_\_\_\_ Прокопьев Н.А.

Заведующий кафедрой

канд. экон. наук, доцент

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_ 2022 г. \_\_\_\_\_\_\_\_\_\_\_\_ Вахитов Г.З.

Казань – 2022

ОГЛАВЛЕНИЕ

[ОГЛАВЛЕНИЕ 2](#_Toc94957680)

[Введение 3](#_Toc94957681)

[Глава 1. Формирование требований 5](#_Toc94957682)

[1.1. Изучение предметной области машинного обучения и нейросетей 5](#_Toc94957683)

[1.2. Анализ существующих подходов к обработке естественного языка 5](#_Toc94957684)

[1.3 Техническое задание 5](#_Toc94957685)

[Глава 2. Проектирование платформы для обучения нейросетевой модели 6](#_Toc94957686)

[2.1. Используемые модели нейронных сетей 6](#_Toc94957687)

[Глава 3. название третьей главы выпускной квалификационной работы 7](#_Toc94957688)

[3.1. Название первого параграфа третьей главы 7](#_Toc94957689)

[Заключение 9](#_Toc94957690)

[Список литературы 10](#_Toc94957691)

Введение

Сегодня автоматическое распознавание речи используется в различных областях жизнедеятельности человека, от умных ассистентов до средств ввода текста голосом. Так как мы свою речь явно не включаем пунктуацию, проблемой большинства алгоритмов распознавания речи является потеря пунктуации. (слайд 3) Это в свою очередь приводит к усложнению чтения полученного текста, по некоторым исследованиям текст без пунктуации сложнее понять, чем текст, в котором 20% слов написаны с опечатками. Таким образом наличие пунктуации критично если результат будут читать люди или будет применяться дальнейшая обработка текста.

(слайд 4) Одним из решений проблемы является применение моделей глубокого обучения.Прорывом в области, безусловно являются предобученные модели на основе трансферного обучения для решения различных задач NLP. Предобученные модели позволяют пропустить самый сложный этап первоначального обучения, и заняться точной настройкой модели (fine-tuning) под конкретную задачу. Одним из таких моделей является BERT.

Существует 2 версии Bert: базовая и расширенная. Базовая версия состоит из 12 слоев кодировщиков, расширенная из 24. Данные кодировщики являются одним из компонентов архитектуры Google Transformers. Данная архитектура раньше использовалась для кодирования последовательности данных во время машинного перевода. Bert использует кодировщики для двунаправленного контекстуализирования слова во вложения, то есть числовые векторы представления слов. Каждый слой пытается закодировать текст в свой контекст. Входной текст представляется в виде токенов. Каждый токен – это отдельное слово или метка. Входной текст представляется в виде токенов. Каждый токен – это отдельное слово или метка. Метки бывают 3 видов: [sep],[cls],[pad]. Первая метка обозначает конец предложения, служит для нужд точной настройки классификации, вторая для обозначения вектора для классификации NSP, а третья – позволяет уравнять входные порции текстов до единой длины. Каждый слой Bert пытается закодировать текст в свой контекст. Задача последнего слоя объединить контексты всех слоев. Таким образом на выходе мы можем получить несколько вложений в зависимости от настройки.

Для точной настройки Bert под конкретную задачу создаётся дополнительный слой нейронной сети, который на вход получает вложение. В дальнейшем происходит обучение финального слоя, сама же NLU модель не переобучается.

Глава 1. Формирование требований

1.1. Изучение предметной области машинного обучения и нейросетей

В последнее время весьма активно обсуждаемым на всех уровнях образования является достаточно новое для отечественной педагогики и образования понятие «культурные (социокультурные) практики в образовательном процессе». Рассмотрим, как трактуется это понятие в педагогической науке.

… … …

1.2. Анализ существующих подходов к обработке естественного языка

Социокультурная практика предполагает последовательное социокультурное освоение мира посредством интеграции общечеловеческого, национального и личностного опыта в процессе образования, сопричастность к национальному, мировому, социокультурному процессу. Именно в силу этого организация социокультурных практик и приобретает социокультурный статус.

С целью обоснования сущностных особенностей социокультурной практики дадим ей характеристику [2].

Во-первых, подчеркнем, что социокультурная практика рассматривается «через человеческое измерение» на основе самостоятельного практического преобразования социокультурного опыта [3-12].

… … …

1.3 Техническое задание

Глава 2. Проектирование платформы для обучения нейросетевой модели

Итак, в предыдущей главе было доказано, что организация социокультурных практик является одной их актуальных образовательных технологий как с точки зрения требований современного социума и культуры к школе, так и в логике ФГОСов (в частности на ступени основной школы).

… … …

2.1. Используемые модели нейронных сетей

Слово «quest» имеет английское происхождение. В переводе с английского языка оно означает «поиск» или «приключение». Именно поиск решения какой-то конкретно поставленной задачи, «путешествие» в эту задачу и составляют основу квест-технологии в образовании.

2.2. Формализация алгоритмов основного анализа текста

Квест-технология – один из форматов игровой технологии. И это очень важно, особенно если говорить о таком сложном с педагогической точки зрения возрасте, как подростковом.

В подростковом возрасте дети постепенно начинают терять интерес к учебе, особенно, если образовательный процесс построен по репродуктивному традиционному типу. Как мы знаем, в подростковом возрасте особую значимость для детей начинают играть такие формы работы, которые позволяют подростку общаться со сверстниками и взрослыми, а общаясь, в совместной деятельности решать значимую как для группы, так и для себя лично проблему.

Глава 3. название третьей главы выпускной квалификационной работы

3.1. Название первого параграфа третьей главы

На сегодняшний день на возможности квест-технологии обращают внимание многие педагоги-практики. Мы в процессе исследования стремились отыскать и проанализировать конкретные варианты реализации интересующей нас образовательной технологии.

Таблица 1. Название таблицы

|  |  |  |  |
| --- | --- | --- | --- |
| Заголовок | Заголовок | Заголовок | Заголовок |
| Текст |  |  |  |
|  | Текст |  |  |
|  |  | Текст |  |
|  |  |  | Текст |

Пользуясь интернет ресурсами мы выяснили, что многие учителя под названием «квест» подразумевают внеурочную игру по станциям или игру, подразумевающую выполнение определённых интеллектуальных заданий на подобие «Своей игры», что несет развлекательную цель, но помимо этого, внеурочный квест имеет возможность привлекать, перемешивать игроков разных классов как для игры, так и для организации. Примерами таких разработок являются: Филологический квест, данный проект нацелен на подведение итогов Недели русского языка и литературы; игра-путешествие "Биологический квест"; Библиотечный проект "Литературная квест-игра «Ночь в библиотеке»" и другие. Еще одним видом внеурочной деятельности является так называемый живой квест, что он из себя представляет? Игры на открытом пространстве, своего рода экскурсии, заранее продуманные, имеющие карту и ряд заданий. Так же на основании проделанной работы можно сказать, что технологию квеста используют для организации самостоятельной образовательной деятельности ученика, примерами являются квест-уроки примером … … .

Рассмотрение и анализ различных конкретных вариантов реализации квест-технологии позволил сделать ряд значимых выводов.

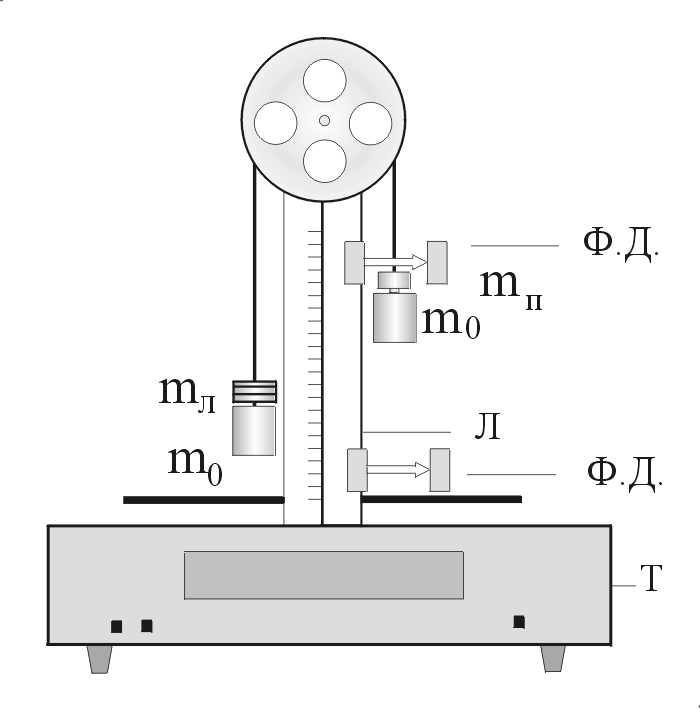


Рисунок 1. Название первого рисунка

Квесты, то есть квестовую технологию, учитывая ее многочисленные возможностями (были подробно рассмотрены выше), используют учителя самых различных предметов (музыки, литературы, географии, физики, математики, истории и т.д.), причем, не только в рамках урочной, но и внеурочной деятельности.

, (1)

К тому же, анализ конкретных вариантов реализации квестовой технологии показал, что конкретные квесты могут иметь разные формы. Действительно, в логике ФГОС учитель вправе выбирать, как именно реализовать технологию квеста в рамках ФГОС.

Заключение

Подведем итоги проведенного выпускного квалификационного исследования и охарактеризуем кратко его основные результаты.

Прежде всего, на основе анализа педагогической литературы по проблеме исследования мы изучили сущность таких понятий как «социокультурные практики» и «квест-технологии».

… … …

Список литературы

* 1. Акулова О. В. Конструирование ситуационных задач для оценки компетентности учащихся: учебно-методическое пособие для педагогов школ / О.В. Акулова, С.А Писарева, Е.В. Пискунова. - СПб.: КАРО, 2008. - 96 с.,
  2. Алексеева Т. Б. Культурологический подход в современном образовании (некоторые теоретические аспекты проблемы) / Т.Б. Алексеева // Мир культуры и культурология Альманах Научно-образовательного культурологического общества России. Русская христианская гуманитарная академия. – СПб.: Научно-образовательное культурологическое общество России, 2013. – с. 200-214.
  3. Алексеева Т. Б. Социокультурные практики в современном образовательном процессе / Т.Б. Алексеева // Развитие педагогической науки в современной России: результаты исследований аспирантских школ (материалы Интернет-конференции). – СПБ.: Изд-во РГПУ им. А.И. Герцена, 2013. – с. 124-127.
  4. Багузина Е. И. Веб-квест технология как дидактическое средство формирования иноязычной коммуникативной компетентности : электронный ресурс : диссертация на соискание ученой степени кандидата педагогических наук: специальность: 13.00.01 / Е.И. Багузина; научный руководитель В.А. Ситаров. – М., 2011. – 238 с. – URL: http://dlib.rsl.ru/01005407532. Режим доступа: для зарегистрированных пользователей на территории виртуальных читальных залов РГБ. (Дата обращения 17.04.2017).
  5. Библиотека материалов для работников школы. Педпортал : сайт. – URL: https://pedportal.net. (Дата обращения 09.05.2017).
  6. Быховский Я. С. Образовательные веб-квесты / Материалы международной конференции "Информационные технологии в образовании. ИТО-99" : электронный ресурс. – URL: http://ito.bitpro.ru/1999. (Дата обращения: 15.04.2017).
  7. ВЕБ - квест "Евгений Онегин" : сайт. – URL: https://sites.google.com/site/vebkvestevgenijonegin/ (Дата обращения 09.08.2017).
  8. Веб-квест по обществознанию "Свобода в деятельности человека" : сайт. – URL: https://sites.google.com/site/vebkvest123/ (Дата обращения 09.05.2017)
  9. Веб-квест "По следам германцев" : сайт. – URL: https://sites.google.com/site/vebkvestposledamgermancev/ (Дата обращения 09.05.2017).
  10. Волкова О. В. Подготовка будущего специалиста к межкульурной коммуникации с использованием технологии веб-квестов : электронный ресурс : диссертация на соискание ученой степени кандидата педагогических наук : специальность: 13.00.08 / О.В. Волкова; научный руководитель И.Ф. Исаев. – Белгород, 2010. – 217 с. –URL: http://dlib.rsl.ru/01004646871. – Режим доступа: для зарегистрированных пользователей на территории виртуальных читальных залов РГБ. (Дата обращения 05.04.2017).
  11. Горлова Н. А. Социокультурные медиаобразовательные проекты на уроках //Справ. зам. дир.-2014.-№2.-С.58.