



СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
SIBERIAN FEDERAL UNIVERSITY



ПРОСПЕКТ СВОБОДНЫЙ- 2017

МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ МОЛОДЫХ УЧЕНЫХ

ЭЛЕКТРОННЫЙ СБОРНИК МАТЕРИАЛОВ
МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ СТУДЕНТОВ,
АСПИРАНТОВ И МОЛОДЫХ УЧЕНЫХ
“ПРОСПЕКТ СВОБОДНЫЙ 2017”
ПОСВЯЩЕННОЙ ГОДУ ЭКОЛОГИИ В РФ

КРАСНОЯРСК, СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

17-21 АПРЕЛЯ 2017 Г.

**Министерство образования и науки Российской Федерации
ФГАОУ ВО «Сибирский федеральный университет»**

Проспект Свободный - 2017

Материалы научной конференции
посвященной Году экологии в Российской Федерации
17-21 апреля 2017 г.

Электронное издание

Красноярск
СФУ
2017 г.

Прикладная лингвистика

КЛАССИФИКАЦИЯ ТЕКСТОВ ПРИ ПОМОЩИ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ И БОТА ДЛЯ TELEGRAM

Кирилов Н.Э.

Научный руководитель ст. преподаватель Николаева Н.В.

Сибирский федеральный университет

Информационные технологии с каждым годом все больше сосредотачиваются на использовании искусственного интеллекта. Отчасти это вызвано активным развитием вычислительной техники. Но во многом интерес к искусственному интеллекту обусловлен теми возможностями, которые он может предложить своим разработчикам. Именно поэтому в каждой крупной IT-компании, таких как Google, Facebook, Microsoft, Amazon, IBM, есть свои отделы, занимающиеся разработками в сфере искусственного интеллекта.

Одной из наиболее перспективных и значительных областей развития искусственного интеллекта является машинная обработка текстов (natural language processing, NLP). Повышенный интерес именно к этой области обусловлен тем, что значительная часть данных, хранящаяся в цифровом виде, это тексты на естественных языках. В данной работе рассматривается сфера машинной обработки текстов, занимающаяся классификацией текстов. Можно выделить следующие основные способы использования классификации текстов:

- 1) Сентимент анализ, позволяющий классифицировать тексты бинарными метками, например: положительное или отрицательное высказывание, спам или полезный email.
- 2) Автоматизированная обработка документов, требующих тэгирувания.
- 3) Построение рекомендательных систем на основе действий пользователей.
- 4) Разработка ПО для работы с библиотечными базами.

На сегодняшний день разработано множество алгоритмов классификации текстов. Все эти алгоритмы варьируются по сложности реализации, числу примеров, требуемых для обучения, а также вычислительным и временным ресурсам, которые необходимо затратить. Помимо этого, к алгоритмам предъявляются требования, чтобы они были устойчивы к переобучению – явлению, когда классификатор хорошо работает на обучающих примерах, но плохо на примерах, которые не участвовали в обучении.

Список из основных алгоритмов классификации выглядит следующим образом:

- 1) Наивный Байесовский классификатор.
- 2) Метод опорных векторов.

- 3) Нейронные сети.
- 4) Метод к ближайших соседей.
- 5) Логистическая регрессия.

В работе реализованы и проанализированы два алгоритма – наивный Байесовский классификатор и метод опорных векторов. Эти алгоритмы были выбраны по причине того, что для их тренировки достаточно небольшого числа примеров. Также стоит отметить, что оба эти алгоритма хорошо работают лишь на линейно разделимых данных, когда признаки одного класса не являются признаками других классов.

Наивный Байесовский классификатор является одним из наиболее распространенных алгоритмов классификации. Алгоритм основан на использовании теоремы Байеса. Популярность алгоритма объясняется возможностью тренировать его на очень небольшом числе примеров. Главным недостатком алгоритма является его низкая точность – обычно она не превышает в лучших случаях 70% на тестовых данных, что обычно дает еще несколько более низкие показатели на реальных данных.

Метод опорных векторов хорошо работает с большим числом признаков и устойчив к перетренировке. Помимо этого, его точность выше, чем у наивного Байесовского классификатора, однако ему нужно больше данных для тренировки.

Наивный Байесовский классификатор

Средняя точность на 127 примерах в соотношении 70 примеров – тренировка, 57 – тестовые составила 60,9%. Средняя точность на 170 примерах в соотношении 100 примеров – тренировка, 70 – тестовые составила 67.4%.

Таблица 1. Результат работы наивного Байесовского классификатора на основе данных, не участвовавших в обучении.

Класс	127 примеров, 20 попыток	170 примеров, 20 попыток
News	66.2%	67.6%
Government	58.5%	63.3%
Hobbies	32.1%	70.1
Класс	127 примеров, 20 попыток	170 примеров, 20 попыток
Reviews	29.7%	79.7
Средняя точность	46.6%	70.2%

В случае для 127 примеров, средняя реальная точность значительно хуже предсказанной (46.6% против 60.9%). Это объясняется тем, что модель подвержена перетренировке, так как в первом наборе текстов преобладал класс «новости». Это приводит к тому, что алгоритм хорошо работает и на тестовых данных, в которых так же содержится

много текстов, помеченных как «новости», однако совершенно не пригоден к использованию на реальных данных.

В случае для 170 примеров из таблицы 1 видно, что работа алгоритма заметно улучшается, как по причине большего числа примеров для обучения, так и по причине того, что эти новые примеры устраняют эффект перетренировки. В результате реальная работа алгоритма получается даже несколько лучше предсказанной (70.2% против 67.4%). Наивысшей точностью угадывания обладает класс «reviews», который угадывается в 8 из 10 случаев. Причину такой высокой точности для этого класса можно объяснить тем, что лексика, присущая текстам данной категории, значительно отличается от остальных, то есть является линейно разделимой. Классы «news» и «government», наоборот, не являются линейно разделимыми – лексика таких текстов может достаточно часто пересекаться, – в результате чего и точность их определения ниже.

Метод опорных векторов

Средняя точность на 127 примерах в соотношении 70 примеров – тренировка, 57 – тестовые составила 32%. Средняя точность на 170 примерах в соотношении 100 примеров – тренировка, 70 – тестовые составила 38.6%.

Таблица 2. Результат работы метода опорных векторов на основе данных, не участвовавших в обучении.

Класс	127 примеров, 20 попыток	170 примеров, 20 попыток
News	39.1%	39.7%
Government	31.2%	32%
Hobbies	23.7%	25.1%
Reviews	19.1%	30.6%
Средняя точность	28.3%	31.9%

Несмотря на то, что метод опорных векторов является более продвинутым относительно наивного Байесовского классификатора, точность его работы много хуже, примерно в 1.7 раза на тестовых данных и в 2.2 раза на реальных данных. Причина этого в том, что число данных, использованных в этой работе, недостаточно для качественной тренировки данного алгоритма.

Бот LingvistBot для Telegram

Бот в данной работе выступает в качестве клиента, позволяющего удобно и быстро работать с обучением и тестированием алгоритмов. В работе реализована работа с наивным Байесовским классификатором, однако это не ограничивает его в тренировке метода опорных векторов (ограничивает только в тестировании).

Стоит отметить, ботом может воспользоваться любой пользователь Telegram, что позволяет тренировать алгоритмы быстрее и делать это людям, которые никогда не имели опыта в программировании и/или машинном обучении. LingvistBot для своей работы использует Telegram BotAPI. Он способен тренироваться по команде /train, выполнять классификацию по команде /test, выводить текущую точность алгоритма по команде /accuracy. Для бота предусмотрен ряд ограничений: размер сообщения для классификации и тренировки должен содержать не менее 500 символов. Интерфейс LingvistBot представлен на рисунке 1.

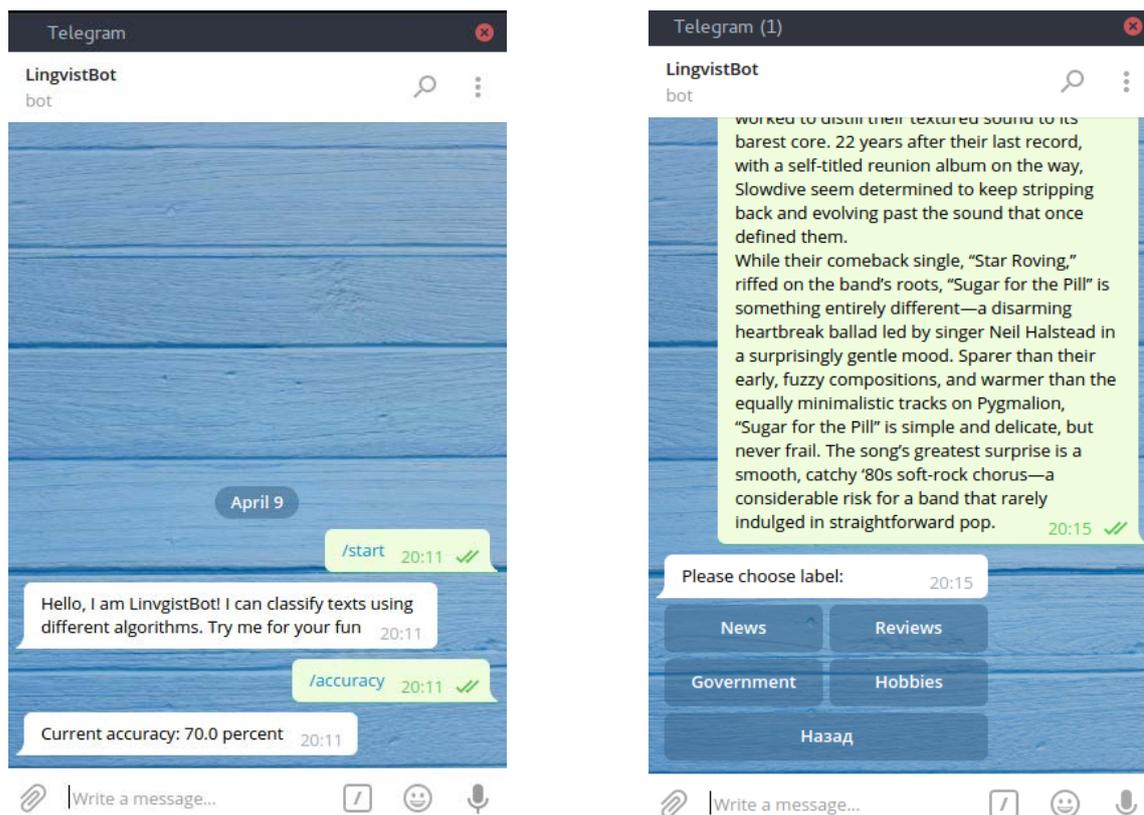


Рис. 1 -Интерфейс LingvistBot

Заключение

Из анализа результатов работы видно, что алгоритмы работают удачно на линейно разделимых данных, однако в случае, когда данные начинают пересекаться (в нашем примере, это новости и политика), алгоритмы чаще дают сбой. Помимо этого, на точность алгоритмов очень сильно влияет число примеров и их сбалансированность, что важно для того, чтобы избежать явления перетренировки.

В конечном счете на точность алгоритмов влияет также то, насколько качественно были подготовлены данные: из данных необходимо убрать знаки препинания, исключить часто встречающиеся слова во всех классах, например, вспомогательные глаголы,

местоимения, союзы, числительные, которые могут иногда значительно влиять на точность работы алгоритмов.

Список использованных источников

1. Bird S., Klein E., Loper E. Natural Language Processing with Python[Текст]: O’Rielly.Inc., 2009. -479с.
2. Thorsten J. Text Categorization with Support Vector Machines: Learning with Many Relevant Features[электронный ресурс]. Режим доступа: https://www.cs.cornell.edu/people/tj/publications/joachims_98a.pdf
3. Joulin A., Grave E., Bojanowski P., Mikolov T. Bag of Tricks for Efficient Text Classification[электронный ресурс]. Режим доступа: <https://arxiv.org/abs/1607.01759v3>
4. Raschka S. Python Machine Learning[Текст:] Packt Publishing Ltd., 2009. -425с.
5. Raschka S. Naive Bayes and Text Classification I[электронный ресурс]. Режим доступа: <https://arxiv.org/abs/1410.5329v3>

ИСПОЛЬЗОВАНИЕ МЕТОДА RSVP ДЛЯ СКОРОСТНОГО ЧТЕНИЯ ТЕКСТОВ

Андруевич С. К.

Научный руководитель канд. пед. наук, доцент Ямских Т. Н.

Сибирский федеральный университет

На чтение различной литературы в среднем человек тратит 3 часа в день или 1080 часов в год. Установлено, что средняя скорость чтения взрослого человека равна 200 слов в минуту. С помощью специальных подготовок люди могут значительно повысить этот показатель и овладеть способностью скорочтения.

Под скорочтением понимают способность чтения в 3 - 4 раза быстрее средней скорости, что примерно составляет 600 слов в минуту. Метод RSVP (rapid solid visual presentation) позволяет читать со скоростью 720 слов в минуту и более. Данный метод быстро нашёл своё применение в скорочтении, а также в других областях.

Технологию RSVP применяет очень малое количество людей, но в наше время данный метод очень актуален из-за развития технологий и увеличения количества портативных устройств, с помощью которых можно использовать программное обеспечение с данной технологией для чтения в любой точке мира. Также он будет полезен в учебной деятельности, так как позволит сократить учебное время на чтение материала учащимися.

Цель работы заключается в поиске эффективного метода обучения скоростного чтения на иностранном языке.

Задачи:

1. Изучить преимущества метода RSVP для обучения скоростному чтению.
2. Провести анкетирование на проверку эффективности использования метода RSVP для обучения скоростному чтению.
3. Выявить возможность применения программного обеспечения, опирающегося на данный метод, для обучения скоростному чтению на иностранном языке (английском).

Преимущество технологии RSVP заключается в том, что слова показываются одно за другим в фиксированной области экрана. При чтении обычным методом глаза человека перемещаются серией чередующихся скачков и остановок. При методе RSVP эти движения перестают быть необходимыми и скорость чтения увеличивается.

Также на скорость восприятия текста влияет режим фокуса, который присутствует в данном методе, при котором человек всегда смотрит на одну выделенную букву. Эта буква

является самым удобным местом в слове, фокусируясь на которую, человек максимально быстро и точно распознаёт слово.

Для того чтобы человек смог быстро читать используя технологию RSVP ему необходимо тренироваться, начинать с малых скоростей, около 250 слов в минуту, и постепенно увеличивать скорость на 50 слов в минуту. За несколько дней возможно достичь скорости 450 слов в минуту, при которой чтение будет комфортным и понимание текста останется на прежнем уровне.

В эксперименте на проверку эффективности метода принимали участие студенты одной из учебных групп ИКИТ СФУ. В ходе выполнения опроса было выявлено, что о данном методе знают 6 человек из 26 опрошенных. После ознакомления с технологией RSVP 16 из 26 опрошенных использовали бы его для постоянного чтения, а 20 из 26 хотели бы видеть его в учебном процессе.

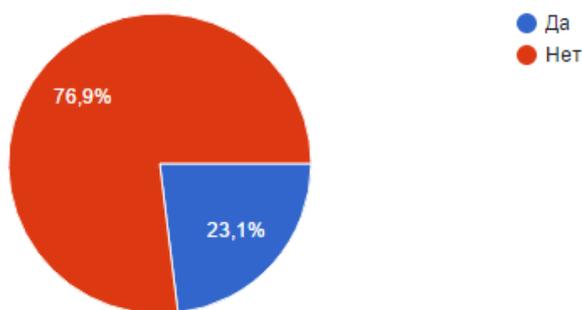


Рис. 1 - Процентное соотношение известности технологии RSVP

Используя технологию RSVP опрошенные проверили с какой максимальной скоростью им будет комфортно читать без предварительной подготовки на двух языках: английском и русском.

Таблица 1 -Таблица скорости чтения опрошенных

Чтение стандартным методом, слов в минуту	Чтение с технологией RSVP (английский язык), слов в минуту	Чтение с технологией RSVP (русский язык), слов в минуту
160	450	550
180	350	450
130	200	250
140	300	350
150	200	300

На данные значения влияют знания языка, а также физиологические и психологические особенности каждого человека.

Программное обеспечение, опирающееся на данный метод, можно использовать в обучении иностранным языкам, как в средних, так и в высших образовательных учреждениях. Для эффективного обучения данным методом знание иностранного языка учащегося должно быть на уровне Pre-intermediate, так как человек должен иметь хороший словарный запас и понимать смысл текста.

Технология RSVP может применяться для быстрого чтения с последующим пересказом. Читая с большой скоростью учащийся не будет успевать запоминать текст, и тогда должен будет выделять главную мысль и пересказывать своими словами, что заставляет его использовать весь свой словарный запас. Таким образом это будет улучшать как письменную, так и разговорную речь учащегося. Также из-за большой скорости чтение текстов будет занимать меньшее количество учебного времени.

Выводы: технология RSVP может позволить достичь большей скорости чтения без явной потери понимания материала в отличие от других способов скорочтения. Данный метод может стать новой ступенью развития учебной деятельности, а именно в изучении языков и других гуманитарных науках.

Список использованных источников

1. Вормсбехер В.Ф., Кабин В.А. *100 страниц в час.*: Кемеровское книжное издательство, 1980, 146 с.
2. Бьюзен Т. *Научите себя думать!*: Попурри, 2004, 192 с.
3. Спенс Р., Витковский М. *Быстрое последовательное визуальное представление: Проектирование для понимания.*: Спрингер, 2004, 110 с. [Spence R., Witkowski M. *Rapid Serial Visual Presentation: Design for Cognition.*: Springer, 2004, 110 p. (in English)].

МОБИЛЬНЫЕ ПРИЛОЖЕНИЯ КАК СРЕДСТВО ОПТИМИЗАЦИИ ИЗУЧЕНИЯ ИНОСТРАННОГО ЯЗЫКА (НА ПРИМЕРЕ АНГЛИЙСКОГО ЯЗЫКА)

Тимошенко Н.С., Ладе А.В.

Научный руководитель ст. преподаватель Ладе А.В.

Сибирский федеральный университет

Введение. В современном информационном обществе становится все более очевидным тот факт, что без знания английского языка невозможно полностью реализовать себя в какой-либо деятельности. Связано это, в первую очередь, с развитием информационных технологий, а также всемирной глобализацией.

В последнее время набирает популярность метод обучения называемый mobile-learning или m-learning. Эта система обучения английскому языку появилась сравнительно недавно - примерно 10 лет назад, и получила широкое распространение в связи с расширением сети интернет, а также развитием мобильных технологий. Поясним, что имеется в виду под m-learning. Электронное обучение с помощью мобильных устройств, не ограниченное и независимое от местоположения учащегося [1]. В настоящий момент больше всего эта система применяется людьми самостоятельно, т.е. факультативно, как дополнительная нагрузка к занятиям в какой-либо языковой школе.

В большинстве научных статей обучение при помощи электронных технологий рассматривается как продуктивное средство в образовании [2], так как предоставляет возможность следить за успеваемостью студентов, их достижениями, индивидуальной программой подготовки. Кроме того, возможность уйти от типовых задач в рамках очного урока считается хорошим способом достижения должного результата. Помимо этого, в русле смешанного обучения преподаватели могут уделять больше внимания индивидуальному подходу к каждому ученику.

Основной процент преподавателей английского языка, занимающихся образовательной деятельностью в университете или школе, стараются применять современные методы обучения и внедрять их в образовательный процесс. Это могут быть как системы электронного обучения, такие как интернет-курсы, так и мультимедийные учебные материалы, видео-подкасты, блоги, вики сервисы и др. [3].

Актуальность данной статьи определяется несколькими факторами. Во-первых, модернизация традиционного обучения ИЯ, т.к. образование в нашем быстро развивающемся мире должно так же быстро адаптироваться и включать в себя новые

системы обучения для подготовки специалистов, уровень которых будет отвечать требованиям современного рынка труда. Следующим фактором является малоизученность влияния мобильных приложений на качество образования. Без должного исследования пользоваться данными технологиями сомнительно, так как данный вид, помимо явных плюсов, может включать в себя и различные минусы, которые впоследствии могут проявить себя.

Исследование. Для более глубокого изучения темы мы провели исследование, которое включало в себя вопросы как общего характера, так и вопросы, направленные на рассмотрение нескольких самых популярных сервисов, позволяющих изучать английский язык при помощи мобильных технологий.

Прежде чем рассматривать мобильные приложения, установим необходимость этого вида обучения в рамках общеобразовательных программ. Для этого воспользуемся методом анкетирования, с помощью средства “Google forms”. В опросе принимал участие 51 человек. Из них 50% - представители Института космических и информационных технологий, 25% опрошенных - студенты Института нефти и газа, 10% респондентов Военно-инженерного института, еще 10% приходится на студентов Института архитектуры и дизайна и оставшиеся 5% - представители Института горного дела, геологии и геотехнологий.

На наш взгляд самым важным является мотивация обучающихся, т.е. их желание изучения английского языка, поэтому мы должны установить, как оценивают свой уровень английского опрашиваемые, какие цели они преследуют, изучая ИЯ.

По данным гистограммы (Рис. 1), отвечая на вопрос о том, как студенты оценивают свой уровень знания английского языка, большинство из них (48%) ответили, что имеют отметку 4, немного меньше (42%) посчитали свои языковые навыки развитыми на оценку “отлично” и только 10% ответили “удовлетворительно”.

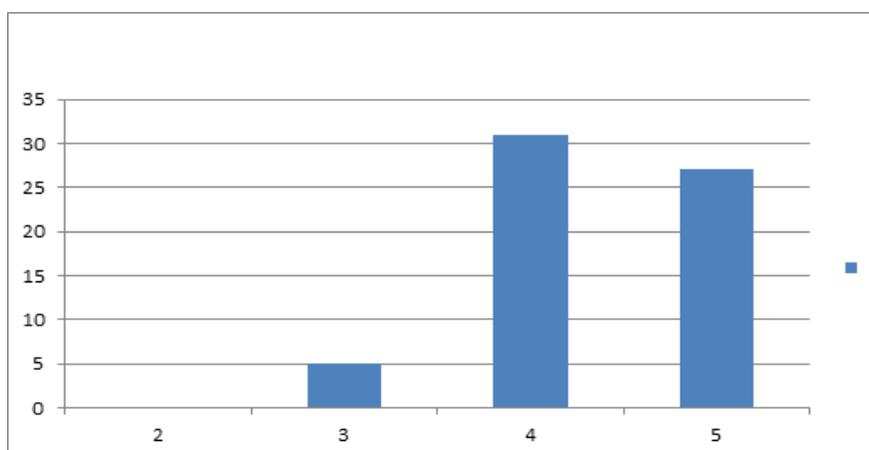


Рис.1 - Оценка обучающимися своего уровня знания английского языка

Далее был задан вопрос о мотивах изучения английского языка, в котором опрашиваемым была дана возможность выбрать несколько вариантов. В результате, наиболее популярным стимулом для овладения английским языком оказалась возможность получать информацию из зарубежных источников, а также необходимость иностранного языка в профессиональной деятельности (Рис. 2). Такие причины, как желание продолжить образование и найти достойную работу за рубежом, интерес к иностранным языкам, а также желание приобщиться к другой культуре, возможность коммуникации с носителями языка и просмотров фильмов в оригинале без субтитров или перевода набрали немного меньше голосов. Наконец, последнее место было отведено удобству использования ИЯ во время путешествия.

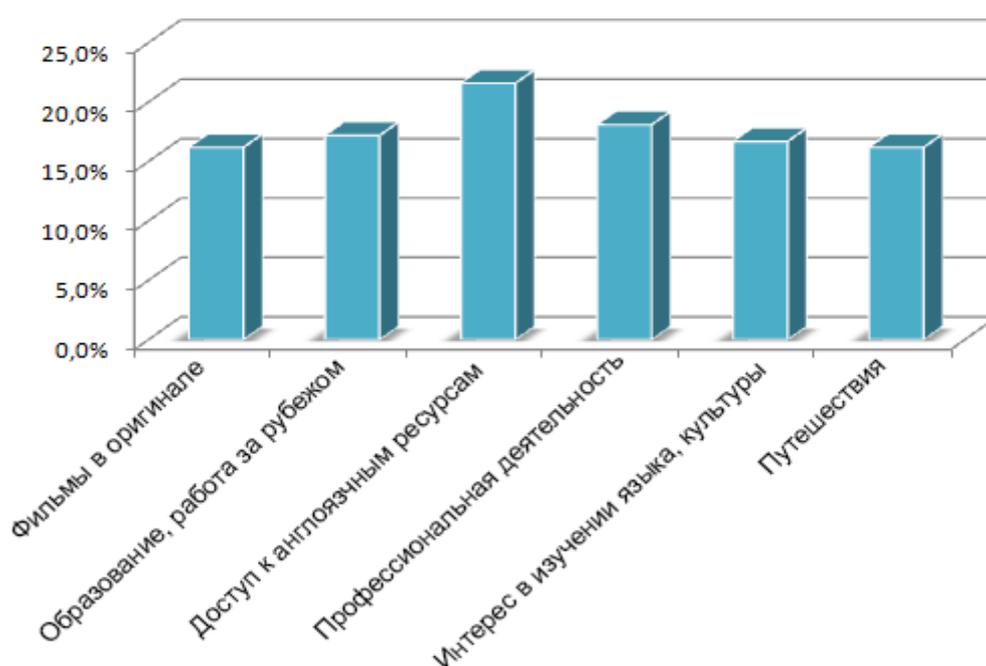


Рис. 2 - Мотивация изучения английского языка

Следующий вопрос касался осведомленности учащихся о мобильных приложениях. Как можно заметить из рисунка 3, подавляющее число опрошенных знают о приложении “Lingualeo”, затем соответственно расположились ответы “Ни о каких”, “Дуолинго”, “Memrise”, пятое место разделили “Busuu” и “Wlingua”, и последнее - “Другое”, в котором не было заполнено поле для ввода текста, т.е. респонденты не указали свой вариант ответа.

О каких мобильных приложения для изучения английского языка вы слышали?

(49 ответов)

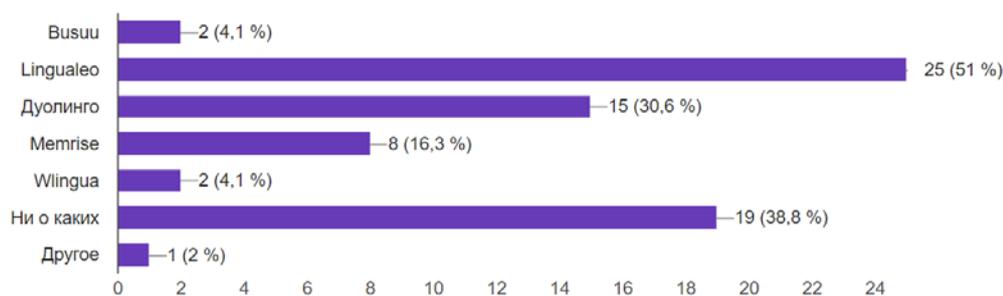


Рис. 3 - Приложения, с которыми знакомы обучающиеся

Следовательно, если большинство людей знают о таких приложениях и, скорее всего, хотя бы раз пользовались одним из них, значит должны как-то оценивать эффективность данных технологий. Отсюда вытекает вопрос о достоинствах и недостатках таких приложений с точки зрения респондентов (рис. 4 и 5). Особое внимание заслужила “Мобильность” данных приложений, т.е. возможность обучения в любом месте и в любое время. Среди следующих преимуществ были отмечены “Сочетание нескольких видов деятельности при обучении (лексика, грамматика, аудирование, чтение)”, “Персонализация (возможность выбора уровня сложности)”, и на последнем месте - “Понятность материала” и “Другое”, которое респонденты оставили не заполненными.

Какие на ваш взгляд плюсы использования мобильных приложений для изучения Ин. языка?

(49 ответов)

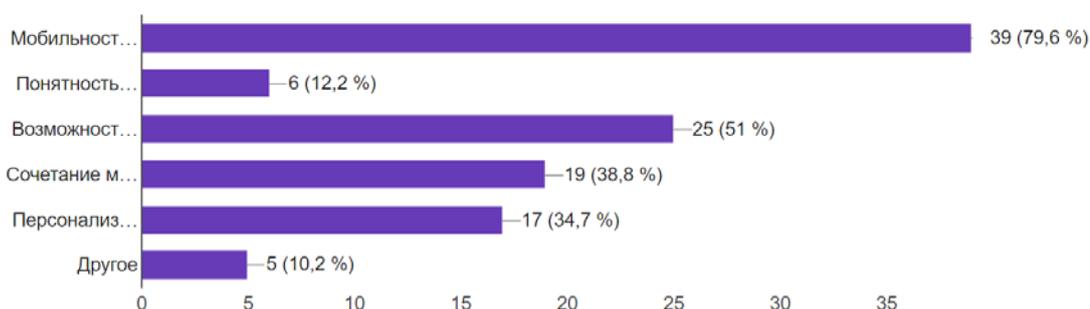


Рис. 4 - Достоинства использования мобильных приложений в изучении ИЯ

Анализируя полученные данные, можно сделать вывод о том, что студенты считают данный вид обучения доступным, но не самым понятным, возможно из-за того, что при таком обучении отсутствует, так называемый, тьютор, который будет проводить контроль за успеваемостью. Это также подтверждается ответами на вопрос о недостатках таких

приложений. Наличие “Рекламы”, “Много платного контента”, “Недостаточности контроля за обучением” являются самыми слабыми местами данных приложений, а варианты ответов “Неинтересно”, “Плохая подача материала” и “Другое”, в котором опрашиваемые так же ничего не написали, оказались невостребованными.



Рис. 5 - Недостатки использования мобильных приложений в изучении ИЯ

Заключение. Исходя из результатов исследования, мы можем сделать вывод о том, что существует необходимость внедрения новых технологий в современное образование для подготовки специалистов высокого уровня. Как показывает проведенная работа “m-learning” является подходящим средством повышения эффективности обучения в рамках курса иностранного языка, так как целевая аудитория (студенты) вполне подготовлена к этому, а преподаватели получают возможность повысить эффективность занятий.

Список использованных источников

1. Академик [электронный ресурс]. Режим доступа: http://normative_en_ru.academic.ru/110470/m-learning.
2. Голицына И.Н., Половникова Н.Л., Мобильное обучение как новая технология в образовании: *Образовательные технологии и общество*, 2011, Том 14 №1, 241-252.
3. Капранчикова К.В., Мобильные технологии в обучении иностранному языку студентов нелингвистических направлений подготовки: *Язык и культура*, 2014, №1(25), 85-94.