

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**АВТОМАТИЗАЦИЯ КОМПЛЕКСНОГО ПОДХОДА К ВЫЯВЛЕНИЮ
SOFT-SKILLS В СЕТЕВЫХ СООБЩЕСТВАХ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

Студентки 2 курса 273 группы
направления 02.04.03 — Математическое обеспечение и администрирование
информационных систем
факультета КНиИТ
Фартушновой Яны Сергеевны

Научный руководитель

к. ф.-м. н., доцент

А. С. Иванов

Заведующий кафедрой

к. ф.-м. н., доцент

С. В. Миронов

Саратов 2023

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Методы анализа профиля в социальных сетях	5
1.1 Влияние активностей в социальных сетях на уровень развития soft-skills	5
1.2 Профайлинг.....	6
1.3 Методы анализа эмоциональной оценки постов в социальных сетях	7
2 Разработка автоматизированной системы поддержки принятия решений	8
2.1 Обработка естественного языка.....	8
2.1.1 Синтаксические преобразования	8
2.1.2 Семантика.....	8
2.1.3 LDA модель	9
2.1.4 Реализация алгоритма тематического моделирования	9
2.2 Модель Word2vec	9
2.2.1 Описание модели word2vec	9
2.2.2 Обучение модели word2vec	10
2.3 Модель рекуррентной нейронной сети.....	11
2.3.1 Рекуррентные нейронные сети	11
2.3.2 Сети LSTM.....	11
2.4 Выделение преобладающих цветов	12
2.4.1 Переход из RGB в CIELAB	13
2.4.2 Реализация алгоритма для выделения доминирующих цветов	13
2.5 Детекция лиц на фотографиях	14
2.6 Маркеры профориентации для оценки личностных и професси- ональных качеств человека	14
2.7 Реализация автоматизированных методов для выделения мета- программ	15
2.8 Разработка интерфейса системы поддержки принятия решений ...	15
ЗАКЛЮЧЕНИЕ	17
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	18

ВВЕДЕНИЕ

С недавнего времени в некоторых зарубежных странах анализируют активность в социальных сетях при приеме на работу кандидата, ссылку на которые он оставляет в своем резюме, так как резюме в настоящее время заполняются по определенному шаблону, в результате чего сложно определить личностные качества и профессиональные способности человека. Психологи проводят множество исследований, изучая модели поведения и речи человека, анализируя отличительные черты на фотографиях и в тексте, которые определяют наличие той или иной черты личности. В результате исследований специалисты определяют маркеры профориентации, на наличие которых проверяют страницы в социальных сетях кандидата и на основе которых делают выводы о его личностных и профессиональных качествах. Такой способ требует наличие подготовленных специалистов в этой области, значительных временных ресурсов, требуемых для тщательной проверки, а также точности статистических данных и объективности анализа. Именно поэтому актуальной и практически значимой задачей является создание автоматизированной системы, позволяющей собирать необходимую информацию из социальных сетей кандидата, требуемую для анализа по психологическим маркерам.

Научная новизна заключается в использовании современных интеллектуальных технологий для анализа информации о кандидате при приеме на работу, ранее проводившемся вручную специалистами в области психологического портретирования, а также в использовании современных психологических исследований специалистов СГУ в области профилирования и портретирования, которые были взяты за основу маркеров профориентации в работе.

Целью настоящей научно-исследовательской работы является разработка методов, требуемых для автоматизированного извлечения и обработки данных из социальной сети «ВКонтакте», для определения личностных и профессиональных качеств кандидата.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Исследование влияния активностей в социальных сетях на показатели soft-skills;
2. Постановка задач для итогового программного продукта;
3. Анализ методов определения тональности текста и разработка алгоритма

- тематического моделирования;
4. Разработка метода для поиска преобладающего цвета среди группы фотографий на основе формулы цветового отличия CIEDE2000;
 5. Реализация алгоритма обнаружения лиц на фотографиях при помощи каскадов Хаара;
 6. Анализ психологических маркеров для оценки профессиональных и личностных качеств человека;
 7. Реализация автоматизированных методов для выделения метапрограмм как инструмента профайлинга;
 8. Реализация методов автоматического извлечения информации из профиля социальной сети «ВКонтакте»;
 9. Реализация автоматизированной системы поддержки принятия решений;
 10. Разработка интерфейса системы с помощью модуля PySide2.

Выпускная квалификационная работа состоит из введения, двух глав, заключения, списка из 62 источников и семи приложений.

В главе «Методы анализа профиля в социальных сетях» рассматривается теоретическая часть используемых терминов и технологий для разработки системы.

В главе «Разработка автоматизированной системы поддержки принятия решений» рассматривается алгоритм предварительной подготовки текстовых данных, описывается класс и методы для определения смысловой нагрузки публикаций, объясняется использование архитектуры Skip-gram модели word2vec для получения векторных представлений слов, а так же выбор для реализации нейронной сети определения тональности текста технологий рекуррентных нейронных сетей и LSTM-слоев, описываются процессы создания и обучения используемых моделей. Также в главе проиллюстрированы и описаны примеры работы созданных и обученных моделей. Также в главе описан класс и методы для определения преобладающих цветов на фотографиях, объясняется использование цветового пространства CIELAB и формулы цветового отличия CIEDE2000. Описана функция определения количества людей на фотографиях, описаны маркеры профориентации и их выявление с помощью программных средств, а также создан и описан интерфейс системы. В главе представлен пример работы приложения на конкретном профиле в социальной сети «ВКонтакте» и описан функционал.

1 Методы анализа профиля в социальных сетях

1.1 Влияние активностей в социальных сетях на уровень развития soft-skills

Soft-skills или личностные компетенции — желательные качества для определенных форм занятости, которые не зависят от приобретенных знаний: они включают в себя здравый смысл, умение общаться с людьми и позитивный гибкий подход.

В ходе психодиагностических исследований специалистами были выявлены зависимости между отдельными видами активностей в социальных сетях и определенным уровнем обладания того или иного soft-skills.

В статье [1] описаны необходимые задачи, которые должен обеспечивать итоговый программный комплекс по автоматическому анализу активностей пользователя в социальных сетях:

1. психодиагностическую деятельность по выявлению soft-skills;
2. диагностику особенностей активности в социальных сетях (количество и частота репостов, цветовая гамма фотографий, использование настоящих фотографий, либо аватарок, количество друзей, эмоциональное содержание постов, обновление информации о пользователе и тому подобное);
3. автоматизированный сбор данных;
4. проведение комплексного корреляционного анализа межличностных особенностей испытуемых (soft-skills) и параметра интернет-активности пользователей в социальных сетях;
5. синтез и обобщение полученных данных;
6. анализ данных кандидатов по их профилям в социальных сетях.

Таким образом, итоговый программный продукт представляет собой систему поддержки принятия решений, с помощью которой специалисты могут делать определенные выводы об уровне необходимых soft-skills у кандидата.

Система поддержки принятия решений определяется как компьютерная автоматизированная система, оказывающая помощь людям, которым необходимо принимать объективные решения в сложных условиях, для полного анализа исследуемой области.

1.2 Профайлинг

Профайлинг — совокупность психологических методов и методик оценки и прогнозирования поведения человека на основе анализа наиболее информативных внешних признаков, а также невербального и вербального поведения.

Для составления психологического профиля человека и оценки его качеств используются основные инструменты профайлинга, основными из которых являются:

1. Метапрограммы - это основные фильтры внимания, которые определяют привычки и поведение человека.
2. Эмоциональное профилирование - анализ эмоциональных реакций человека, определение ведущих воздействий и влияния на него.
3. Пирамида ценностей и актуальных целей - в зависимости от того, какие у человека ценности и цели определяется его поведение. Наличие схожих пунктов важно в построении взаимоотношений между людьми.
4. Характер и его маски - определение действительного характера человека, который часто скрывается за привычными окружающим масками.
5. Лингвистический и пресуппозиционный анализ речи - результат такого анализа определяет образ мышления человека, его жизненные установки, таланты и, наоборот, уязвимые стороны. Однако, это является достаточно кропотливой работой.

В настоящей работе в качестве инструмента профайлинга будут использоваться метапрограммы, описанные в первом томе сборника Алексея Филатова «Профайлинг. Как научиться разбираться в людях и прогнозировать их поведение» [2].

В современном понимании метапрограммы – это базовые фильтры восприятия человека и фокусирования его внимания, которые применяются ко всему спектру опыта и определяют образ мышления, актуальные ценностные ориентации, личностные качества, стереотипное поведение, привычки и, соответственно, принципы и образ жизни.

Для каждой метапрограммы существует два или более альтернативных сценария, которые формируют развитие определенных качеств личности, стереотипов, ценностей. Инструментарий оценки индивидуальных ценностей и личностных качеств по метапрограммам имеет убедительное и доказательное основание.

1.3 Методы анализа эмоциональной оценки постов в социальных сетях

Существует множество методов определения тональностей текстов.

Классификация методов определения эмоциональной окраски текстов:

1. Первая группа основана на эмотивной лексике в тексте.
2. Методы второй группы различаются наличием или же отсутствием обучающей выборки корпуса данных.

2.1. Обучение без учителя.

Обучение без учителя определяется как раздел машинного обучения, в котором определяются закономерности и взаимосвязи между объектами из некоторой неразмеченной выборки данных.

2.2. Обучение с учителем.

Обучение с учителем предполагает наличие определенного набора выборок в качестве входных данных, где ожидаемые выходные данные, или метки, уже известны и находятся в соответствии с элементами выборок.

Эффективными методами обучения с учителем в задачах определения эмоциональной окраски считаются: метод опорных векторов, наивный Байесовский классификатор и метод максимальной энтропии.

В статье [3] также приведено формальное сравнение методов определения эмоциональной окраски текста с применением обучения с учителем. Авторы статьи приходят к выводу, что при анализе точности и полноты лучшие результаты вне зависимости от способов представления векторов, содержащих признаки дает метод опорных векторов. Анализ также показал, что результаты использования наивного Байесовского классификатора и метода максимальной энтропии практически одинаковые, однако первый значительно проще в реализации, что дает ему преимущество в использовании.

2 Разработка автоматизированной системы поддержки принятия решений

2.1 Обработка естественного языка

Современные компьютерные системы могут понимать естественные языки благодаря базовой технологии, называемой NLP (Natural Language Processing — обработка естественного языка). Эта технология способна обрабатывать естественный язык в качестве входных данных и выполнять ряд специальных операций, связанных непосредственно с целями тематико-нейролингвистического характера:

1. Анализ тональности текста;
2. Классификация текстовых документов;
3. Принятие решения о действии на основании определенного заявления;
4. Извлечение намерений.

2.1.1 Синтаксические преобразования

Для успешной обработки текстовых данных программа должна получать на вход предварительно подготовленный и обработанный корпус текстов. Предварительная обработка состоит из последовательного выполнения следующих действий:

1. Разбиение текста на предложения.
2. Удаление знаков препинания и токенизация.
3. Лемматизация слов или стемминг.
4. Удаление «стоп-слов».

Итогом текущего раздела является функция предварительной подготовки текста.

2.1.2 Семантика

Вторым важным разделом анализа текста является семантический подход. Семантика относится к анализу тематической и смысловой нагрузки текста. Семантический анализ является одним из сложных аспектов обработки естественного языка, который еще не полностью решен. Он включает в себя применение компьютерных алгоритмов, которые позволяют понять значение и интерпретацию слов, а так же то, каким образом структурированы предложения.

Обработка естественного языка играет важную роль в поддержке взаимодействия между человеком и машиной и помогает упростить жизнь во многих сферах. Применяя и группируя алгоритмы из семантического и синтаксического подходов, можно с помощью использования простейших алгоритмов машинного обучения выискивать теги(основные ключевые слова текста), которые и будут являться смысловой нагрузкой в тексте.

2.1.3 LDA модель

Тематическое моделирование — прием машинного обучения без учителя для определения тем в коллекции документах. Тематическое моделирование определяет темы, к которым относится каждый из документов коллекции и какие слова (термины) образуют каждую тему.

Модель латентного размещения Дирихле (LDA) позволяет реализовать тематическое моделирование коллекции текстов. После извлечения тем из корпуса текстов каждый из документов представляется как набор вероятностей извлеченных тем. Дополнительно темы так же представляют собой вероятностное распределение через слова. Уникальная особенность моделей LDA состоит в том что темы не обязательно должны быть различными и слова могут встречаться в нескольких темах; благодаря этому учитывается некоторая нечеткость, присущая естественному языку.

2.1.4 Реализация алгоритма тематического моделирования

Автоматическое извлечение темы из большого потока документов — это одна из основных задач обработки естественного языка. Для данной задачи на языке *Python* предусмотрен ряд библиотек для решения различного рода задач в *NLP*.

Основными библиотеками для автоматического извлечения темы будут служить библиотеки *gensim*, *nltk*, *pymorphy2*.

Итогом текущего раздела является класс и методы для извлечения смысловой нагрузки текста.

2.2 Модель Word2vec

2.2.1 Описание модели word2vec

После предварительной подготовки необходимо перевести полученные слова естественного языка в форму, которую способна принять на вход и

проанализировать нейронная сеть. Для этой цели выберем модель для анализа семантики естественных языков word2vec, разработанную в 2013 году группой исследователей компании Google под руководством Томаша Миколова.

Word2vec — набор алгоритмов для расчета векторных представлений слов, получающий на вход корпус текста и создающий на выходе набор векторов слов.

Принцип работы данной модели основан на идее о том, что лингвистические единицы, встречающиеся в похожих контекстах, являются семантически близкими, или означают похожие вещи.

В word2vec содержится два типа алгоритмов: на основе моделей CBOW (Continuous Bag of Words) и Skip-gram. Алгоритмы, основанные на CBOW пытаются предсказать слово по заданному контексту, используя простую модель мешка слов с учетом четырех ближайших соседей термина (два предыдущих и два последующих слова), и для хороших результатов применяются при выборке более ста миллионов слов. В архитектуре Skip-gram предсказывается контекст на основе заданного слова, используя последовательности k-skip-n-gram длины n , в которых элементы находятся на расстоянии друг от друга не более k , и такая модель может показывать хорошие результаты на выборке менее ста миллионов слов. Для реализации модели была выбрана архитектура Skip-gram, которая, хоть и обладает меньшей скоростью обучения, но лучше работает на относительно маленьких выборках и учитывает редкие слова.

2.2.2 Обучение модели word2vec

Корпус текстов предварительно обрабатывается, а затем подается на вход модели, затем поочередно выполняются следующие действия:

1. Происходит считывание корпуса текстов, затем подсчитывается частота встречаемости в нем каждого слова.
2. Полученные слова образуют словарь, которые сортируются по вычисленной частоте встречаемости. Словарь обладает определенным размером, поэтому из него исключаются редкие слова.
3. Обработка происходит по субпредложению (это может быть как предложение исходного текста, так и целый абзац), модель проходит по ним окном определенного размера. Размер окна — максимальная длина между текущим словом и словом, которое предсказывается. По умолчанию размер окна равен 5, но оптимальным считается значение 10.

4. Данные, находящиеся в текущем окне попадают на вход нейронной сети прямого распространения.

В качестве модели word2vec была выбрана реализация, предоставляемая библиотекой Gensim, так как она предоставляет удобные инструменты для многопоточной обработки больших корпусов текстов, что позволяет производить обучение быстрее.

Gensim предоставляет возможность выполнять загрузку и обучение модели встроенными средствами. В качестве обучающей выборки был взят корпус коротких текстов на русском языке на основе постов twitter Юлии Рубцовой [4].

2.3 Модель рекуррентной нейронной сети

2.3.1 Рекуррентные нейронные сети

Рекуррентные нейронные сети работают с последовательностями и списками. Они используют уже полученную информацию для решения будущих задач. В некоторых случаях для решения поставленной задачи требуются только последние из полученных данных, например, для предсказания следующего слова в словосочетании, в этом случае нескольких последних слов будет достаточно, чтобы сделать вывод. Рекуррентные нейронные сети справляются с задачами, в которых расстояние между необходимой информацией и местом ее использования для решения следующей задачи небольшое.

Но существуют ситуации, которые требуют больше контекстных данных, тогда расстояние между предыдущей информацией и местом, в котором она требуется, увеличивается. С учетом увеличения такого расстояния связь между информацией теряется. Теоретически с этим можно справиться, если исследовать все зависимости и правильно подобрать необходимые параметры сети, но существуют исследования, доказывающие фундаментальные ограничения рекуррентных нейронных сетей, а в следствие и неспособность решать такие задачи на практике. [5]

2.3.2 Сети LSTM

LSTM (long short-term memory, долгая краткосрочная память) — тип рекуррентной нейронной сети, который способен обучаться долгосрочным зависимостям.

LSTM созданы специальным образом для решения проблемы долгосрочных зависимостей. Они запоминают информацию в течение длительных периодов времени, поэтому их практически не нужно обучать.

Архитектура всех рекуррентных сетей имеет форму цепочки, состоящей из повторяющихся в цикле модулей сети. Если в стандартных нейронных сетях такие повторяющиеся модули имеют простую структуру, к примеру, только один слой, то в LSTM повторяющийся слой может состоять из четырех связанных слоев. [6]

Именно LSTM будет использована в данной работе, но для начала нужно осуществить обработку данных. В качестве тренировочной и тестовой выборок будет так же использован корпус коротких текстов на русском языке на основе постов twitter Юлии Рубцовой. Текстовый корпус нужно обработать с помощью описанной выше функции предобработки текста.

Полученный обработанный корпус необходимо разделить на тренировочную и тестовую выборки, а затем разбить текст на токены, преобразовать этот текст в последовательность чисел в соответствии с полученным словарем для подачи на вход нейронной сети. Далее для того, чтобы на вход модели позитивные и негативные данные подавались попеременно, перемешаем их.

Далее необходимо загрузить предварительно обученную модель word2vec для того, чтобы инициализировать матрицу Embedding слоя весами, полученными при обучении модели word2vec.

После того, как необходимые данные подготовлены и обработаны, нужно создать модель рекуррентной нейронной сети с использованием рекуррентного слоя LSTM.

Результатом текущего раздела является обученная модель LSTM-сети для определения тональности текста.

2.4 Выделение преобладающих цветов

Алгоритмизация восприятия цветов человеческим глазом — одна из самых сложных задач в изучении компьютерного зрения. Существует большое количество цветовых пространств, учитывающих различные особенности зрения человека. В ходе изучения проблемы было выяснено, что цветовое пространство LAB и формула цветового отличия CIEDE2000 являются наиболее приближенными к человеческому глазу.

Цветовое пространство CIELAB представляет собой аппаратно-незави-

симулю модель «стандартного наблюдателя». Цвета, которые она определяет, не связаны с каким-либо конкретным устройством.

2.4.1 Переход из RGB в CIELAB

Для перехода из цветового пространства RGB в цветовое пространство CIELAB нет прямой формулы. Для этого сначала необходимо перейти из RGB в пространство XYZ, а потом из XYZ уже в пространство CIELAB.

Код преобразования был проверен с помощью сайта для дизайнеров, где один и тот же цвет представлен во всех возможных цветовых пространствах. В связи с вычислительными потерями неточность составляет отклонение от реального цвета в CIELAB на значение 0.001.

2.4.2 Реализация алгоритма для выделения доминирующих цветов

Для выделения доминирующих цветов был выбран метод кластеризации *k*-means — метод *k*-средних. Идея метода заключается в том, чтобы при кластеризации данных минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров. На первом этапе выбираются случайным образом начальные точки (центры масс) и вычисляется принадлежность каждого элемента к тому или иному центру. Затем на каждой итерации выполнения алгоритма происходит пересчет центров масс — до тех пор, пока алгоритм не сойдется [7].

Расстояние между двумя точками в пространстве CIE XYZ не имеет отношения к относительной близости этих цветов. Именно поэтому для пространства CIELAB используется не Евклидова метрика, а CIEDE2000.

В специально предоставленных для исследования психологических маркерах для анализа профиля в социальной сети были заданы основные цвета: белый, черный, желтый, красный, синий, зеленый, фиолетовый и коричневый, — после выделения цветов три оттенка должны быть соотнесены к одному из данных цветов, для этого требовалось точное нахождение расстояние между цветами с помощью формулы CIEDE2000.

Результатом текущего раздела является класс и методы для определения доминирующих цветов на фотографии

2.5 Детекция лиц на фотографиях

Обнаружение объектов на фотографиях — это технология, связанная с компьютерным зрением и обработкой изображений.

Алгоритм обнаружения осуществляет поиск экземпляров семантических объектов выделенного класса на фотографиях или видеозаписях. Для выполнения поставленной задачи необходимо осуществить обнаружение лиц на фотографиях.

Будем использовать для выполнения задачи библиотеку OpenCV и Каскады Хаара. [8]

Каскад Хаара — метод, основанный на машинном обучении, для обучения которого используются положительные и отрицательные выборки. Позитивные изображения содержат объекты, которые необходимо классифицировать, отрицательные, соответственно, — не содержат.

Итогом текущего раздела является функция определения количества лиц на фотографии.

2.6 Маркеры профорientации для оценки личностных и профессиональных качеств человека

Психологами из Саратовского государственного университета были представлены следующие лингвистические маркеры для анализа информации из социальных сетей:

1. Использование в тексте названий кинофильмов, музыкальных композиций, компьютерных игр различных жанров;
2. Частотность употребления местоимения «я» на странице пользователя;
3. Косвенное обоснование своего мнения после цитируемого текста:
 - а) Имена известных авторов,
 - б) Названия известных журналов;
4. Частотность употребления имени пользователя другими членами исследуемой социальной сети;
5. Частота публикаций пользователя;
6. Частота использования в публикациях прилагательных и глаголов;
7. Статус доступности профиля;
8. Использование репостов;
9. Использование URL-адресов;

10. Количество отметок «нравится» на публикациях;
11. Эмоциональная окраска излагаемого текста;
12. Реакция сообщества (эмоциональная окраска комментариев к изложенному сообщению);
13. Самоидентификация рода деятельности или профессии;
14. Смысловая направленность репостов;
15. Цветовая гамма фотографий.

Для извлечения требуемой информации из социальной сети ВКонтакте будем использовать модуль vk [9] для языка программирования Python, разработка которого организована на сервисе GitHub.

2.7 Реализация автоматизированных методов для выделения метапрограмм

В своих социальных сетях пользователи оставляют значительное количество информации о себе и своих интересах, эта информация, а также способ ее подачи и отличительные особенности ведения страницы, может рассказать многое о личности анализируемого пользователя.

Приведем перечень метапрограмм для определения их у профиля в социальной сети «ВКонтакте».

1. Мотивация ОТ и Мотивация К
2. Активный и Рефлексия
3. Внутренняя референция и Внешняя референция
4. Процедуры и Возможности
5. Общее и Детали
6. Сходство и Различия
7. Сам, Другой и Система
8. Ассоциация и Диссоциация

Результатом текущего раздела являются функции определения метапрограммной принадлежности профиля.

2.8 Разработка интерфейса системы поддержки принятия решений

Интерфейс проекта будет создан с использованием модуля PySide2, который является проектом для создания привязки Qt к языку программирования Python. Он направлен на реализацию поддержки всех возможностей Qt, в том числе QtQuick и QtMobility.

PySide2 предоставляет доступ к программе QtDesigner, в которой можно создать собственный интерфейс. Для разработки системы необходимо пять форм:

1. Первая форма нужна для ввода ключа доступа «ВКонтакте», требуемого для работы с vk_api.
2. Вторая форма необходима для ввода ссылки на профиль, или его идентификатора, и количества последних публикаций, анализ которых требуется провести.
3. В третьей форме будут отображены имя и фамилия пользователя, а также кнопки выбора: «Рассчитать принадлежность к метапрограммам» и «Узнать информацию со страницы пользователя», при нажатии каждой из них будет отображаться отдельное окно с соответствующей информацией.
4. В четвертой форме будут отображаться результаты, полученные в результате анализа профиля «ВКонтакте» пользователя по предоставленным психологическим маркерам профориентации.
5. В пятой форме будут отображаться результаты выявления метапрограмм в процентном соотношении, описанных в книге Алексея Филатова [2], на странице в социальной сети «ВКонтакте».

ЗАКЛЮЧЕНИЕ

В ходе выполнения настоящей работы были выполнены следующие задачи:

1. Изучено влияние активностей в социальных сетях на показатели soft-skills, приведены исследования, доказывающие применимость анализа soft-skills по активностям в социальных сетях;
2. Поставлены задачи для итогового программного продукта, принято решение взять за основу проекта систему поддержки принятия решений, изучены возможности систем поддержки принятия решений;
3. Проанализированы методы определения тональности текста, сделаны выводы об эффективности их использования, рассмотрены достоинства и недостатки основных методов.
4. Проанализированы существующие алгоритмы для тематического моделирования и выбран метод LDA для реализации программы на языке Python;
5. Разработан алгоритм, позволяющий осуществлять тематическое моделирование текстов публикаций из социальных сетей;
6. Разработан метод для поиска преобладающего цвета среди группы фотографий на основе формулы цветового отличия CIEDE2000;
7. Изучены психологические маркеры для оценки профессиональных и личностных качеств человека;
8. Реализован алгоритм обнаружения лиц на фотографиях профиля в социальной сети «ВКонтакте» при помощи каскадов Хаара и OpenCV;
9. Разработаны методы для автоматизированного выделения метапрограмм как инструмента профайлинга по информации о пользователе из социальной сети «ВКонтакте»;
10. Реализованы методы для автоматического извлечения информации из профиля социальной сети «ВКонтакте»;
11. Реализована автоматизированная система поддержки принятия решений для выявления soft-skills и метапрограмм по профилю пользователя в социальной сети «ВКонтакте»;
12. Разработан интерфейс системы поддержки принятия решений с помощью модуля PySide2.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Иванов А. С., Фартушнова Я. С. Шапкина Д. И. Киселева О. Н. Ярошенко Е. И. Воронин А. В.* Автоматизация комплексного подхода к выявлению soft-skills в сетевых сообществах / Фартушнова Я. С. Шапкина Д. И. Киселева О. Н. Ярошенко Е. И. Воронин А. В. Иванов, А. С. — 2021.
- 2 *Филатов, А. В.* Профайлинг. Как научиться разбираться в людях и прогнозировать их поведение / А. В. Филатов. — Москва: Перо, 2016. — С. 417.
- 3 *Н. Ю. Рязанова, К. М. Сперцян.* Сравнительный анализ методов определения эмоциональной окраски сообщений в социальных сетях с применением обучения с учителем / К. М. Сперцян Н. Ю. Рязанова // *Новые информационные технологии в автоматизированных системах.* — 2018. — no. 21. — Pp. 417–423.
- 4 Корпус коротких текстов на русском языке на основе постов твиттер RuTweetCorp Ю. Рубцовой [Электронный ресурс]. — URL: <https://study.mokoron.com/> (Дата обращения 08.05.2023). Загл. с экр. Яз. рус.
- 5 LSTM — нейронная сеть с долгой краткосрочной памятью [Электронный ресурс]. — URL: <https://neurohive.io/ru/osnovy-data-science/lstm-nejronnaja-set/> (Дата обращения 14.05.2023). Загл. с экр. Яз. рус.
- 6 Знакомство с архитектурой LSTM-сетей [Электронный ресурс]. — URL: <http://datareview.info/article/znamomstvo-s-arhitekturoy-lstm-setey/> (Дата обращения 16.05.2023). Загл. с экр. Яз. рус.
- 7 Colour metric [Электронный ресурс]. — URL: <https://www.comrphase.com/cmetric.htm> (Дата обращения 21.04.2022). Загл. с экр. Яз. англ.
- 8 Поиск лиц на изображении с использованием OpenCV в Python [Электронный ресурс]. — URL: <https://waksoft.susu.ru/2021/04/02/raspoznawanie-licz-s-ispolzovaniem-opencv-v-python/> (Дата обращения 10.12.2022). Загл. с экр. Яз. рус.
- 9 API|VK для разработчиков [Электронный ресурс]. — URL: <https://dev.vk.com/reference> (Дата обращения 08.04.2022). Загл. с экр. Яз. англ.