***ФГБОУ ВО «Кубанский государственный аграрный университет имени И.Т. Трубилина»,***

***Российская Федерация***

***Апасов Максим Вадимович, ПИ2102***

***macsim.apasov@yandex.ru***

**137 Теория информации, данные и знания. Лабораторная 7. 2020-10-13**

**Заголовок:** Лабораторная работа по системе "Эйдос": Спектральный АСК-анализ изображений художников и задачи для студентов

**Резюме**

Данная сессия представляет собой седьмую лабораторную работу для группы ИТ-2003, проведенную 13 октября 2020 года профессором Вовченко Евгением Вениаминовичем. Основная тема – освоение спектрального АСК-анализа изображений с использованием интеллектуальной системы "Эйдос".

**1. Введение и контекст:**
Занятие начинается с приветствия и уточнения старосты группы. Устанавливается контекст: 6-я пара, 7-я лабораторная работа по дисциплине "Теория информации, данные, знания".

**2. Спектральный АСК-анализ изображений:**
Инструктор вводит новую тему – спектральный анализ изображений как метод исследования предметной области через модель.

* **Данные:** Студентам предоставляется ссылка на скачивание набора изображений – картин известных художников (Айвазовский, Да Винчи, Куинджи, Рерих, Шишкин). Объясняется структура именования файлов (Художник-Номер) и необходимость поместить их в папку indata. Рекомендуется использовать программу ACDC для приведения изображений к формату JPG/BMP и стандартному размеру (ширина 200 пикселей).
* **Работа в системе "Эйдос":** Демонстрируется пошаговый процесс анализа в системе:
	+ Запуск системы и проверка актуальности версии.
	+ Импорт данных изображений через режим 2.3.2 (выбор подрежима оцифровки по пикселям и спектру).
	+ Формализация предметной области: создание базы данных изображений, загрузка файлов, создание классификационных (имена художников) и описательных (спектральные интервалы) шкал.
	+ Настройка параметров анализа (пиксели, спектр, цвета) с рекомендацией использовать значения по умолчанию.
	+ Синтез и верификация моделей (ИНС-2, ИНС-3) для оценки их достоверности.
	+ Визуализация результатов в режиме 4.7: отображение спектров отдельных картин и усредненных спектров по классам (художникам). Интерпретация спектров для выявления характерных цветовых палитр каждого художника.
	+ Демонстрация результатов распознавания (идентификации) картин по их спектрам с использованием построенной модели.
	+ Визуализация структуры модели: матрица модели, нейроны, когнитивная диаграмма, дендрограмма для анализа сходства/различия художников по цветовым спектрам.

**3. Технические вопросы и устранение неполадок:**
Обсуждается проблема одного из студентов с запуском системы "Эйдос" после обновления. Даются рекомендации по проверке настроек Windows (согласно инструкции), запуску исполняемого модуля напрямую и отмечается потенциальная нестабильность Windows 10 по сравнению с Windows 7.

**4. Следующие шаги: Разработка собственных приложений:**

* **Задача для студентов:** Основная дальнейшая работа – самостоятельная разработка собственного приложения в системе "Эйдос". Студенты должны выбрать тему, найти или подготовить набор данных (текстовых, табличных или графических), используя ресурсы вроде UCI или Kaggle, либо собственные данные.
* **Методология:** Необходимо следовать инструкции по разработке, использовать шаблоны описания приложений.
* **Оценка:** Итоговая оценка будет зависеть от размещения работающего приложения в облаке "Эйдос", качества его описания, а также от публикации результатов исследования на платформах ResearchGate и РИНЦ (для получения высших баллов).

**5. Заключение:**
Инструктор подчеркивает, что студенты освоили базовые навыки работы с различными типами данных в "Эйдос" и готовы к самостоятельной проектной работе. Занятие завершается предложением задать вопросы.

**Детальная расшифровка текста**

**1. Введение и контекст**

Здравствуйте, ребята.
Здравствуйте.
Да, здравствуйте, здравствуйте.

Так, кто у нас староста?
Группа 2003.
М?
Староста заходит, сейчас он зайдет.

Алло, здравствуйте, зашел.
Здравствуйте.
Так, Волков, да?
Да, Волков, староста ИТ-2003.
Мм.
То есть сегодня что, у нас уже было занятие с вами?
Да, было сегодня.
Обалдеть.
Два занятия у нас в один день.
Понятно.

Ну, значит, давайте тогда продолжим. Я под запись скажу, что у нас сегодня 13 октября 2020 года. Шестая пара с 17:20 по 18:50. Седьмая лабораторная работа с группой ИТ-2003 по дисциплине Теория информации, данные, знания.

И у нас осталось рассмотреть несколько еще аспектов исследования модели, исследования предметной области путем исследования модели.

**2. Спектральный АСК-анализ изображений**

Но я предлагаю вам сейчас рассмотреть такой вопрос. Это мы рассмотрим, но на другом, на другой лабораторной работе, которую я называю спектральный АСК-анализ изображений.

То есть, я вам говорил же, да, ребята, что система Eidos позволяет обрабатывать изображения.

Да, ребята, еще я забыл сказать, что занятие проводит профессор Вовченко Евгений Вениаминович. Это же все под запись. То есть вы должны понимать, что эта информация вот сигнальная такая, в самом начале занятия, она нужна. Потому что потом позже уже не поймешь, что за занятие, кто его проводил, зачем его проводил. Хотя и так непонятно тоже, если честно. Даже когда и говоришь.

Вот. Значит, сейчас я установлю приложение.

Значит, ребята, для того чтобы у вас это работало, вам нужно скачать...

**2.1. Данные и подготовка**

Хм. Странно, странно.

Сейчас я вам дам ссылку. Скачаете файлы с этой, с этой, по этой ссылке.

Так. Я поделился изображением, ребята? Еще, наверное, нет, да? Совместный доступ.
Нет, у вас только видео, ну, то есть вебка.
А, да, я вижу, ага. Вот, теперь вижу, что...
Вот сейчас поделились.
Да, поделился, ага. Просто я упустил это из виду.

Значит, теперь смотрим, ребята. Значит, вот у нас Яндекс Диск. Видеозанятия. Спектральный анализ. Поделиться, копировать ссылку. Вот, я вам даю ссылку в чат. Значит, вы сейчас эти изображения, которые там есть, себе скачаете. Систему не обязательно скачивать. Потому что вы же обновление провели? Да? Вы же обновляли систему, ребят?
Не все успели еще.
Обновляя. Поэтому, значит, и достаточно.

Вот скачиваете эти изображения. И если вы обновляли систему, у вас все должно сейчас работать.

Теперь, куда мы записываем эти файлы, ребята? В папочку indata записываем файлы графические.

Что представляют собой эти файлы? Это картины известных художников, выдающихся: Айвазовского, Да Винчи, Куинджи (кстати, Куинджи - учитель Рериха), Рериха Николая Константиновича, Шишкина. Вот. То есть это по несколько картин каждого художника здесь приведено.

Что позволяет делать система Eidos? Она позволяет вводить изображения в систему и анализировать их. Она позволяет это делать разными способами: и по контурам, и по пикселям, и по пикселям, и по контурам, и только, в общем, в разных сочетаниях все это. И, значит, эти режимы анализа изображений по пикселям я их разрабатывал, и они в основном в тексте программы реализованы. А по контурам есть модуль, который Дима Бандык разрабатывал, и есть модуль, который я разрабатывал по контурам.

Так вот, что позволяет делать система Eidos с изображениями? Она позволяет формировать обобщенные образы изображений определенных категорий.

Значит, обратите внимание, как имена этих файлов, какие у них имена. У них имя состоит из двух частей. Первая часть повторяется - это фамилия художника, потом тире и номер. И вот мы можем таких номеров, то есть картин разных, то есть одного и того же художника, сделать не три, а 10 там, 20 и так далее. И самих самих художников можем сделать не пять там, а 10, 20, 30.

Таким образом, эти вот все вот картины одного художника будут использованы для формирования обобщенного образа картин этого художника.

Теперь, э-э, давайте выйдем на мой сайт. Второй пункт. И во втором пункте здесь вот есть инструкция для учащихся по разработке собственных Eidos, облачных Eidos приложений. И вот здесь у нас есть, ребята, я его немножко меняю этот файл инструкции. Вот. И сегодня, наверное, постараюсь изменить его тоже.

Вот. Значит, здесь есть шаблонное описание. Вот после этой лабораторной работы, которую сейчас мы пройдем, мы начнем заниматься непосредственно разработкой собственных приложений. То есть вы будете выбирать тему, вот, и так далее.

Здесь есть шаблонное описание приложений разной направленности. И есть шаблон описания, который вот соответствует тому, что мы сейчас будем делать. Я показываю вам ссылочку на статью.

После обновления что не запускается? Как как это не запускается? Вы запускали... Ну я понимаю, что программа. Марк, вы запускали Start Eidos или саму систему запускали? И что, раньше запускалась эта система у вас на компьютере?

Попробуйте сейчас вот с этого скачать облако, где вот я давал вам ссылочку на файлы, оттуда попробуйте скачать. Это для меня новость.

Ладно. Значит, э-э, этот вот файлик можно открыть статьи и поглядывать, что там такое. Значит, там статья из двух частей состоит. Первая - это описано коротко вообще о том, что такое системно-когнитивный анализ. И потом численный пример использования спектрального автоматизированного системно-когнитивного анализа, который сейчас мы и будем изучать. И здесь видно, что те же самые картинки мы используем. Вот они. И, в общем, как они представлены эти картинки, видно, видите? Вот. Как мы будем их вводить?

**2.2. Работа в системе "Эйдос"**

Значит, вводить... Давайте теперь... Ребят, вы мне отвечайте, пожалуйста, голосом, потому что лабораторная работа. Значит, вы мне отвечайте голосом, чтобы я постоянно не листал эти странички, экраны. И вы тоже, Марк, тоже голосом мне говорите, получилось, не получилось.

Значит, в системе Eidos есть много различных программных интерфейсов, то есть программ, предназначенных для ввода данных из внешних источников данных. Это все эти интерфейсы собраны в режиме 2.3.2. Это импорт данных из текстовых файлов, универсальный программный интерфейс импорта данных в систему из табличных данных, еще один такой же есть, только другие стандарты экселевской таблицы. Оцифровка изображений по внешним контурам, оцифровка изображений по всем пикселям и спектру.

Вот давайте этот запускаем этот режим оцифровка по всем пикселям и спектру. Значит, если, ребята, у вас что-то не получается, вы не успеваете там или что-то не работает, вы мне голосом говорите, хорошо?

И смотрите, я буду постараться вам помочь. Значит, у нас появляется такое окошко, где выбор двух вариантов: либо мы формализацию предметной области проводим, вот, либо создание распознаваемой выборки.

Значит, формализация предметной области включает разработку классификационных и описательных шкал градаций, кодирование исходных данных с их использованием и получение обучающей выборки. А распознаваемая выборка создается с уже имеющимися классификационными и описательными шкалами градациями. И создается не обучающая выборка, а распознаваемая. Сейчас мы выбираем первый пункт.

Появляется четыре пунктика у нас, ребята. Окошко с четырьмя пунктиками. Первый пункт - пересоздать базу данных изображений, то есть ну стереть её. Потом загрузить изображения. Вот они загружаются эти изображения в эту базу данных. Просматривать нет никакой необходимости, но можно и просмотреть. Вот. И потом выбираем четвертый пункт - создать базу данных для стандартного программного интерфейса 2.3.2.3, которым мы еще не пользовались.

Сейчас мы, когда выбрали этот четвертый пункт, появилось у нас окошко, где можно задать, смотрите, сколько, что мы можем анализировать изображение только по пикселям, только по спектру, только по пикселям и спектру. Сколько цветовых, сколько цветов в спектре, спектральных диапазонов сколько задать. Истинный цвет черный, истинный белый цвет, фон не учитывать, формировать файл, полная формализация предметной области.

Значит, ничего здесь менять не надо. Просто я вам показываю, что есть разные варианты. Здесь уже такие варианты подобраны параметров, которые разумно дают результаты. Можно экспериментировать, проверять разные варианты.

Начинается анализ изображений, ребята, создание базы исходных данных для программного интерфейса 2.3.2.3. И, значит, сейчас начнется анализ изображений сам.

Это будет занимать некоторое время у нас. Вообще обработка изображений - это трудоемкий процесс в вычислительном отношении. Вот сейчас вот первое, сейчас второе проанализировано изображение, третье.

Еще это существенно зависит от размеров изображений. Система Eidos может обрабатывать изображения, ну она может вообще-то любого размера обрабатывать разумного. Иногда возникают ошибки при очень больших изображениях, там, ну, допустим, 2К. Вот. Но, значит, если там, допустим, 640 на 480, она обрабатывает такие изображения. Но я вам это не советую делать, потому что времени много будет это занимать. Лучше все изображения привести к определенному стандарту. Ну, во-первых, нужно знать, что она обрабатывает только JPEG и BMP изображения. То есть нужно их преобразовать в JPEG, видимо. Для этого есть смысл использовать программу ACDC. Очень хорошая программа, удобная. Позволяет пакетное преобразование осуществлять.

И обратите внимание, ребята, появилось сообщение, в котором дается понять пользователю, что теперь делать дальше. Создана база данных для стандартного программного интерфейса 2.3.2.3. Вот. Теперь нужно запустить этот интерфейс с параметрами по умолчанию. После этого надо запустить синтез и верификацию модели в режиме 3.5, а также в режиме 4.7 визуализацию спектров объектов и классов. Можно также смотреть все выходные формы во всех режимах, как обычно, как мы это делали.

Значит, теперь мы должны запустить интерфейс 2.3.2.3 с параметрами по умолчанию. Эти параметры, они сформированы автоматически предыдущим режимом. Просто мы нажимаем О'кей. Всё, осуществлен процесс формализации предметной области.

Смотрим, что у нас получилось. Классификационные шкалы - это название из имён файлов взята та часть, которая до тире, до номера реализации. То есть это фамилии художников, пять фамилий художников. Вот. Дальше смотрим 2.2. Значит, здесь у нас шкала одна - спектральные интервалы, то есть спектральные диапазоны. И здесь у нас мы задали 35 спектральных диапазонов. Все эти спектральные диапазоны есть. Причем для удобства пользователя, они прямо изображены эти цвета. Там где вот код градации, там прямо не просто этот код написан, а еще прямо эта градация изображена.

И можем посмотреть обучающую выборку. Здесь у нас в этой обучающей выборке 16 картин. И каждая картина относится к одному классу. А сбоку мы видим, ребята, сбоку мы видим коды каждого пикселя. То есть все пиксели здесь перечислены. И мы видим код каждого пикселя в соответствии со справочниками спектральных диапазонов. Прямо конкретно указан код пикселя, в какой спектральный диапазон он попадает.

Теперь мы создаем модели и проверяем их на достоверность.

Вот, у меня это заняло 12 секунд. Мне интересно, сколько это у вас заняло, ребята. Давайте покажите мне, сколько это заняло у вас в чате. Просто напишите там 12, 10, там 8, 30 там.

Вот, так вот ACDC программа, она позволяет пакетное преобразование имён осуществить. То есть можно вот так вот набрать часть имени, потом тире, и она сама пронумерует файлы в папочке. Создаст, то есть можно так вот очень удобно переименовывать файлы, ну, допустим, картины художников. Второй момент - она позволяет быстро преобразовать формат графических файлов. То есть можно из любого формата быстро получить JPEG формат. Третье - она позволяет сделать ресайз, изменение размеров файлов. Я вам советую делать файлы размером 200 по ширине. Ширина файлов 200. Вот. Хотя можно где-то до 450, я вам говорил. До 400. Ну, лучше делать около 200, чтобы было реально быстрее можно было это обработать.

Ребят, что-то я не вижу никакой информации от вас, обратной связи нет у меня. Это же лабораторная работа, вы должны делать то же самое, ребят. Пишите, сколько у вас заняло времени синтез модели. Если вы не осуществляли синтез модели, пишите минус тогда. Я буду знать, что вы не выполняете лабораторную работу.

У Шамиля очень хороший компьютер. Максим, а что у тебя за ошибка вышла? Ошибка исполнения? У тебя ноутбук или что? Какая у тебя там ошибка? Что за ошибка?

Ладно, значит, давайте, я так понял, что дело туго. Значит, теперь заходим в режим 4.7. О чём было написано, ребята, и вот в этом хелпе, который появился после окончания ввода изображений в систему. Значит, здесь есть помощь, ребята. Мы в этой помощи видим, что нужно делать для обработки этих изображений. Какие режимы нужно выполнять, в каком порядке. Вот. Теперь, и что это вообще позволяет сделать обработка изображений по спектрам. Вот здесь вот у нас всё это написано. То есть я прямо взял и написал всё это.

Теперь, я хочу вам дать ссылочку одну ещё на подборку публикаций моих по обработке изображений. Я так понял, что никто ничего не делает. Те, кто делают, у них не особо-то получается.

Вот. Что-то я вообще из системы вышел. Это зря.

Заходим в режим 4.7. Запускаем... Вот здесь у нас есть два меню: 4.8 и 4.7. Заходим в меню 4.7 и изображаем, выводим изображение спектров объектов. Ну у нас там один почему-то файл не воспринимается, не знаю почему даже, если честно. И другие файлы мы видим, они вот сейчас на экране меняются и выводятся справа спектры.

Что такое спектр? Это информация о том, каких пикселей сколько в этом изображении. Какой спектральный диапазон попадают эти пиксели. И в процентах показано от общего, ну, все сумма пикселей - это 100%, и каждый цвет показан, сколько процентов пикселей такого цвета от всех пикселей, которые есть в изображении. Мы видим, что она так довольно-таки неплохо это определяет, что вот синий цвет представлен там, допустим, мы это видим. Тут зелёные представлены хорошо цвета, коричневые. Шишкин. Вот. И, в общем, вот у нас получилась такая картина.

Теперь мы здесь видим сообщение, что все эти изображения записаны в папку со спектрами inx spectrum pix. Вот. То есть спектральные характеристики прямо самих этих вот изображений. Да, кстати, когда мы создали приложение, то здесь вот в режиме 1.3 не мешало бы написать, что это спектральный АСК-анализ изображений. Вот. Эта информация появляется во всех формах выходных. Графических, по крайней мере.

И теперь выводим изображение... Да, ребят, сначала смотрим, какая модель более достоверна оказалась. Смотрим, насколько эти модели вообще достоверны. Значит, я показываю вам частотное распределение согласно критерию Ван Рейсбергена. Показываю частотное распределение по обобщению этого критерия Ван Рейсбергена, которое я предложил, мультиклассовое нечёткое обобщение. И по-другому критерию тоже показываю. Ну, в общем, что могу сказать? Получилось, что, в общем-то, ну, наверное, вот имеет смысл модель INX2 использовать с критерием сумма знаний. Вот.

Ну давайте сейчас спектры посмотрим классов, а потом уже решим задачу идентификации. Вот, модели INX2. Значит, что мы здесь видим? Мы слева видим спектр класса соответствующего, а пунктиром обозначено среднее по всей выборке. То есть используется классическая модель АСК-анализа.

Сейчас я вам... Ну и была информация о том, что все эти спектры записаны в папочку. Значит, я вам показываю модели, ребят, смотрите на модели. Вот модель INX2. Здесь мы видим слева спектральные диапазоны, а справа видим, насколько эти спектральные диапазоны характерны, не характерны для тех для картин тех или иных художников. Видите, вот Шишкин, Рерих, Куинджи, и мы видим плюсы и минусы. Минус - это значит, эти пиксели есть такого цвета, но не характерны, то есть они встречаются у этого художника реже, чем в среднем по всей выборке. А какие-то встречаются чаще. Вот у Да Винчи, например, красные, коричневые очень часто встречаются. Вот. Смотрим матрицу абсолютных частот. Здесь у нас большие сравнительные числа, потому что в этих картинах пикселей всё-таки довольно много. Вот. И вот мы видим, как эти пиксели распределены по спектральным диапазонам и по художникам. В каком количестве какие пиксели у тех или иных художников в картинах этих художников встречаются. Понятно? В общем, понятно.

И смотрим саму модель. Мы здесь видим, какие пиксели характерны, какие не характерны для тех или иных художников. Посмотрим, может быть, когнитивные функции. Я про когнитивные функции вам рассказывал, ребят? На прошлом занятии, нет? Нет, не рассказывал, да?

Сейчас посмотрим. Да, это не прокатит. Жаль, красиво было бы.

Вот, ребят, значит, теперь давайте посмотрим на спектры классов. Значит, смотрите, ребята, значит, мы видели с вами, что в этой вот матрице модели есть у нас, ну, скажем так, вероятности, вероятность встречи какого-то пикселя у объектов какого-то класса и средняя вероятность встречи по всей выборке. И эти вероятности сравниваются методами, которые я вам показывал, раньше объяснял подробно. И получается модель, описывающая, насколько тот или иной пиксель характерен для того или иного класса, какую несёт информацию о принадлежности классу. То есть это сравнение условных относительных частот с безусловными относительными частотами, даёт такой результат. Есть разные способы их сравнения: хи-квадрат, например, там и другие. И мера Харкевича, коэффициент возврата инвестиций. Все они используются для этих целей.

И вот мы смотрим, что у нас получилось в результате, ребята. Вот у нас распределение пикселей в классе Айвазовский на основе полученной на основе его картин, которые есть там в обучающей выборке. А пунктир, ребята, это в среднем, вероятность встречи этого цвета в среднем по всей выборке. Мы видим, что некоторые цвета у Айвазовского встречаются чаще, чем в среднем по всей выборке. Это характерные для него картин цвета. А некоторые встречаются реже, чем по всей выборке. Это цвета, хотя у него и встречаются, но они для него менее характерны, чем для других художников. Мы видим, что у Айвазовского наиболее характерными для него являются зелёно-голубые цвета. А вот красно-коричневые там и какие-то там запредельные пурпурные цвета, фиолетовые, пурпурные, у него, хотя и есть, но представлены в меньшей степени, чем у других художников.

Теперь смотрим Да Винчи. Мы видим, что среднее распределение числа пикселей по цветам, то есть средний спектр по всем по всем картинам, вот так можно сказать, он тот же самый пунктир. А у Да Винчи у него намного больше, чем в среднем представлены коричневые цвета, а также фиолетовые, сиреневые. А зелёные у него, хотя и есть, но они относительно меньше.

Теперь смотрим на Куинджи. У него зелёные, коричневые цвета представлены чаще, чем у других художников. Здесь вот слева мы это видим. А голубые, фиолетовые, пурпурные у него реже встречаются, чем у других художников.

Мы видим, ребята, что Рерих, он намного меньше у него красно-коричневых цветов, чем в среднем по всей выборке наших картин. А фиолетовые и пурпурные у него встречаются синие, фиолетовые, пурпурные чаще. Ну я так думаю, это связано с тем, что он рисовал в основном горы, а в горах там ультрафиолет, всё освещено в таких вот горы, в таких вот голубых и фиолетовых тонах.

И Шишкин. У Шишкина вполне ожидаемо красно, то есть это зелёные и коричневые цвета представлены, а синий и пурпурный у него в меньшем количестве представлены.

Вот. Теперь на основе этой информации система может осуществлять идентификацию конкретных картин, относить их к тем или иным художникам по их спектру. Для этого мы сейчас сделали модель INX2 текущей и решили задачу идентификации. Смотрим результаты решения задачи идентификации. Вот у нас эти картины. Здесь надо сделать имена короткие. Что я сделал длинные, я не понимаю. Надо сделать короткие имена вот здесь.

Вот. И мы что видим? Что Айвазовский относится к Айвазовскому. Айвазовский 2 тоже к... тоже к Айвазовскому относится, да? Ну и к Рериху ещё второй. Вот. Да Винчи 1 - Да Винчи. Да Винчи там 2 - тоже Да Винчи. Да Винчи 3 - тоже. Куинджи 1 - она ошиблась, но хотя тоже его указала, но на Шишкина больше показалось похожее. Куинджи 2 - тоже на Шишкина больше оказалось. Значит, Рерих верно, верно, верно. Рерих все верно. Шишкин тоже все верно. Ну то есть, можно сказать так, что система может идентифицировать картины, относить их к тем или иным художникам по спектрам этих картин.

Теперь давайте посмотрим на SWOT-анализ. То есть мы можем сказать, что для Айвазовского, допустим, наиболее характерными являются вот эти цвета - сине-зелёный спектр. А те вот нехарактерными. Вот мы видим слева, что сине-зелёный спектр характерен для Айвазовского, а зелёно-коричневый, красный не характерен. Ну мы это и так знаем, в общем-то. Вот Да Винчи. Вот для него характерны цвета вот такие красно-фиолетовые, пурпурные. А зелёно-коричневые не характерны. И вот так мы видим по каждому художнику. Шишкин. Понятно, что у него, для него характерны зелёные и коричневые цвета, а голубые, фиолетовые, пурпурные не характерны. Они используются для рисования для в живописи Рериха, Айвазовского. А вот эти вот - это характерно именно для изображения деревьев, листвы. Вот. Ну, в общем, вот так вот, в общем, мы видим, что для кого характерно.

Теперь мы можем сравнить этих художников по спектрам их картин и нарисовать когнитивную диаграмму сходства-различия этих художников по спектрам картин. И мы видим, что какие-то художники в какой-то степени сходны. Вот Рерих, например, получился немножко похож на Айвазовского, потому что такая голубая гамма там есть. А Куинджи похож на Шишкина, а между собой они не похожи. Это же можно увидеть в форме дендрограммы агломеративной. Здесь мы то же самое видим: Куинджи на Шишкина похож немножко, но больше, чем другие художники друг на друга. Вообще надо сказать, что они довольно сильно друг от друга отличаются. Вот хотя Айвазовский и Рерих они похожи, но они похожи не очень сильно, понимаете? Но если сравнить с другими, то вот они Куинджи и Шишкин больше всего похожи друг на друга, потом похожи поменьше Айвазовский и Рерих по цветовой гамме. И Да Винчи он там где-то посерединке между ними всеми уникальный. График расстояний межкластерных.

И посмотрим теперь, насколько цвета различные несут сходную или отличающуюся друг от друга смысл, насколько они сходны, отличаются по их смыслу. А их смысл в том, что они несут информацию о принадлежности, непринадлежности картины с этим цветом к тому или иному художнику. Вот мы видим, что сходными являются цвета красные, фиолетовые, сиреневые, фиолетовые. Вот. И вот мы видим, какие из них похожи. И вот, допустим, эти зелёные похожи друг на друга. Вот. Коричневые, зеленоватые, коричневатые похожи, зеленоватые, вот они похожи друг на друга. Вот как они объединяются. Конструкты, то есть кластеры, конструкт. На одном полюсе конструкта вот эти цвета находятся, на другом - зелёные. Это по этим художникам мы получили смысл этих цветов. Как они связаны друг с другом по смыслу. И график межкластерных расстояний.

Ну, я сейчас вам, ребята, бегло описал некоторые возможности системы Eidos по анализу изображений по их цвету. Что, в общем-то, наверное, тоже довольно интересно.

**3. Технические вопросы и устранение неполадок**

Так, теперь смотрим 8:50, 18:50. Это ещё у нас 40 минут времени. Поэтому давайте сейчас начнём рассматривать новый вопрос. Я так понял, что с таблицами мы поработали, в том числе с методикой риэлторской оценки. Ребят, вы мне подтвердите это сейчас. Действительно так или нет? Второе, мы поработали с текстами. Вы ознакомились с тем, что можно в системе Eidos обрабатывать тексты, да? Правильно? Следующее, мы ознакомились с тем, что можно обрабатывать изображения.

**4. Следующие шаги: Разработка собственных приложений**

Таким образом, вы сейчас готовы решать задачи обработки текстов, табличных данных и изображений в системе Eidos. И на учебном таком уровне, и на таком более уже имеющем практическое значение уровне, которое мы с вами изучили на основе методики риэлторской оценки автомобилей с пробегом, которую мы создали в системе Eidos. Было дело, ребят? Подтверждаете? Было или не было? Что вы там молчите? Будьте более активными.

Сейчас попробовал переустановить всю систему на компьютер, но программа всё равно не запускается. Марк, а у тебя она работала вообще на компьютере, нет? Система-то? Работала. Ну, всё странно всё это. Честно сказать, не очень даже и понятно. То работало, то перестало работать. А какие-то сообщения выдаёт оно, нет? Ты запускаешь самоисполняемый модуль системы, не Start Eidos, да, Марк? Исполняемый модуль. Ну и что, она просто не работает или что-нибудь сообщает? У нас должны какие-то ошибки выдаёт. Она просто не открывается. Вообще не запускается, что ли? То есть как бы зависает, да? Хм. Ну у меня она работает, ты же видишь. У меня эта версия. Вы пробовали из облака её скачать, да? Вот скачивали, и она вот не работает, да? Или вы полностью скачали инсталляцию? Полностью тоже. Хм. Да. А папка как называется, где находится система? Она в корневом каталоге находится? То есть на корневом диске C, да? Ну... А вот настройки системы Windows, вот эти вот вы делали, которые я сейчас вам переслал в чат? Раньше делали, да, наверное? У вас всё работало. Ну тогда там читать не очень удобно. Я сейчас ссылочку дам вам. И вы там откроете и почитаете в отформатированном виде. Потому что чат этот вот портит форматирование. А! Ну ясно. Значит, э-э... Ну, ребят, честно сказать, я не знаю, с чем связано. Значит, я могу вам сказать моё впечатление. В этой системе Windows 10 операционке всё работает хуже гораздо, чем в семёрке. То есть она очень капризная эта система. Там э-э очень переусердствовали с защитой информации. Прямо вот, ну, чрезмерная, понимаете, там параноидальная какая-то система. Очень капризная. Значит, я вам послал ссылочку, ребята, где можно скачать систему Eidos, да? А немножко вниз пролистываете, пониже. Ну там буквально страницу, там страницу. И там написано некоторые замечания по поводу операционной системы. Вот. И касаются эти замечания прежде всего Windows 10. Вот этот текст ищите там. И там пункты, 10 пунктов. Если вы их выполните, очень большая вероятность, что всё заработает. Ну я могу сказать так, что вот из моего опыта я знаю, что у многих она просто работает и всё, на десятке. То есть они вот просто установили, запустили и всё, вопросов особых не возникло. В Windows 7, ребята, вообще таких проблем нет. Вообще нет проблем подобных. То есть там всё работает. То есть это очень хорошая, стабильная версия системы Windows, удачная, я считаю. А вот восьмёрка там всякие Миллениумы, десятка - это неудачная версия. У них бывают периодически то удачные, то неудачные версии. Возможно, потом они сделают опять удачную, какую-нибудь там двадцатку. У них так бывает, они экспериментируют. И у них с переменным успехом это получается. Я помню, вот до семёрки была, было много между семёркой и XP разных версий. Вот никто не смог превзойти XP. XP была великолепная система, Windows XP. Потом много было версий, которые получились очень неудачными. И и восьмёрка там была, и Миллениум был там, всё там только не было. И что это вообще люди попользовались и бросили их. А потом появилась семёрка. Вот семёрка, ребята, это можно сказать так, доделанная XP с учётом, ну, более новых реалий, скажем так, по аппаратному обеспечению. Она очень удачная получилась. Никакого, все, ни у кого нет никаких вопросов по этому поводу. Потом появилась восьмёрка и десятка. Восьмёрка вообще, так сказать, попробовали и бросили. Десятка стали пользоваться, но просто, я бы сказал так, под давлением обстоятельств, потому что там драйвера под новое оборудование и так далее. Вот, а старое перестали поддерживать операционные системы, перестали под них писать драйвера. И новое компьютеры, они просто не работают под другими версиями системы операционными. То есть могут быть проблемы с этим. Вот. Поэтому, конечно, если вы переустановите семёрку, то просто будет работать и всё. И не только система Eidos, но и другие системы просто будут работать и не будут патать вам нервы. Вот. А в десятке можно попробовать. Вы сначала не переустанавливаете, а попробуйте вот эти настройки сделать, которые я вот рекомендовал вам. Может быть, и будет нормально работать, потому что я знаю по опыту, что у многих она так прекрасно работает на десятке. Правда, почему-то Start Eidos не очень хорошо работает, который обновляет систему, проверяет её корректность её исполняемого модуля.

**4.1. Задача для студентов: выбор темы и данных**

Значит, что теперь вам делать вот в данный момент, что вам делать? Значит, я советую вам эту инструкцию почитать и перейти, сейчас мы уже три пункта прошли этой инструкции. Вот эти три пункта мы прошли. Ну, первый пункт - это вообще информация о технологии Eidos, второе - это вот лабораторные работы мы изучили базовые, основные по основным направлениям. Теперь мы должны понять, что мы можем делать приложения, которые удовлетворяют нескольким требованиям. Первое - они не противоречат действующему законодательству Российской Федерации. Второе - они не противоречат морально-этическим нормам. Третье - эти приложения ещё не сделаны. Вот мы смотрим здесь вот, здесь уже имеющиеся приложения находятся. Их 208. Ну здесь кэш думает, что их 207. Вот, это уже имеющиеся приложения. Любое из этих приложений можно почитать описание по-любому из них. Вот. Значит, если вы увидите, ребята, что описание ущербное, то есть сейчас вы видите, что в системе лучше можно это всё сделать, вот, и лучше получить результаты, лучше оформить, то это можно сделать. То есть можно даже брать имеющиеся уже работы, если они устарели, если их описание устарело, и вы можете лучше это сделать и дополнительные возможности исследования реализовать, то тогда можно это вполне сделать.

Потом есть ещё вот такие моменты. Вот я, допустим, беру, набираю слово телефон. А у нас есть два исследования с телефонами: исследование активности мобильных телефонов в городе с использованием, да? А также вот у нас есть анализ спроса на мобильные телефоны. И всё. А больше у нас про телефоны ничего нет. Можно вполне и ещё работу провести про телефоны. И про Винчестеры, или про видеокарты, или ещё про что угодно. Много работ нет, ребята.

Значит, ничего сдавать не надо. На следующей паре, значит, вы должны определить свою тему. Сейчас я как раз про это рассказываю, что мы должны делать, в общем-то, дальше. Дальше вы должны выбрать тему своей работы. Я хочу сказать, что довольно много тем у студентов связано с играми. Они ещё не очень выросли, помнят, как они играли, некоторые ещё продолжают и до сих пор играть. И некоторые игры, то есть некоторые приложения интеллектуальные, созданные студентами, связаны с анализом игр, персонажей, ситуаций. Есть довольно такие интересные, сильные работы в этой области. Потом я могу вам сказать, что игра, она ведь является в каком-то смысле моделью реальности. То есть если можно сделать анализ ситуаций в игре, точно так же можно сделать анализ ситуаций в жизни, аналогично.

Значит, дальше вы должны что сделать? Значит, это вот мы прошли первые три пункта. Вот теперь мы выполняем четвёртый пункт. Четвёртый пункт заключается в том, что вы должны определиться с темой своей работы. Значит, вы должны поискать, какие есть наборы данных для машинного обучения. Их очень много сейчас. Может быть, вы там найдёте что-то для вас интересное. Вы должны также знать, что информация для набора данных для машинного обучения, которую можно использовать в этих вот ваших работах, их можно взять на сайте UCI. Здесь их довольно много. На любой клацаем, там у нас прямо вот наборы данных, читать данные, это вот это и так далее. Вот я уже считал данные. Вот. Понимаете? То есть фактически большое количество здесь есть данных на сайте UCI. Это древний сайт, который существовал ещё аж, сейчас я вам скажу когда, вы даже удивитесь. Значит, чтобы сказать, когда он существовал, я сейчас выйду на страничку, где мы скачиваем систему Eidos и ищем UCI. Ребята, вот в 2003 году я написал статью. Это было 17 лет назад. Методика использования репозитария UCI для оценки качества математических моделей систем искусственного интеллекта. Вот. Давно это было, давно.

Потом времена изменились, был создан сайт Kaggle. Сначала это было самостоятельное явление, то есть независимое от других фирм. А не так давно, ну, может, год назад, этот портал, можно сказать, Kaggle, его купила фирма Google. Почему? Он стал очень заметным, очень интересным. В чём преимущество Kaggle по сравнению с UCI? Значит, есть два преимущества. Первое - у него на сайте Kaggle информация представлена в более стандартном для системы Eidos виде, в виде CSV файлов. Это текстовые файлы, форматированные, в которых поля отделены запятой или точкой с запятой, которые в Excel вводятся на раз-два-три, грубо говоря. То есть берём, значит, вот, то есть вкладка Данные, вот данных из текстовых файлов и выбираем разделитель, и потом О'кей, и всё, и у нас файл в Экселе, понимаете? То есть это очень удобно. На UCI там может быть всякие заумные форматы, которые надо там самому менять. Но в моей статье об этом написано, как это меняется.

Потом, смотрите, ребят, сейчас я вам покажу ещё статьи с этим UCI. Чтобы вы видели, что не только в 2003 году, но и позже были написаны статьи про этот репозитарий. Вот, смотрите, здесь их было семь штук, видите? Первая статья, вторая статья, третья статья, четвёртая, пятая, шестая, седьмая. Понимаете? То есть довольно много статей вот здесь вот у меня, в которых можно найти ссылочку на UCI.

Значит, если вы выйдете сюда и зададите запрос UCI, то вы найдёте семь статей, которые основаны на обработке данных с этого портала. Вот. Но они такие, я сказал, немножко более заумный формат данных. Хотя для человека, который немножко владеет Экселем, никакой проблемы нет, чтобы эти данные преобразовать в нужный формат. На Kaggle проще. Но на Kaggle есть иногда простые приложения, где один файл исходных данных, они называют это тренировочная выборка. Вот. И один, скажем, тестовый файл, и всё. А бывает там, допустим, пять или шесть, там, 10 файлов исходных данных, которые связаны отношениями один ко многим друг с другом там, ещё в каких-то отношениях находятся. То есть иногда надо делать программу для того, чтобы это всё объединить там и так далее. Я думаю, нет смысла это делать, ребят. То есть никакой необходимости в этом нет. То есть ищите приложения попроще и делайте.

Теперь следующий момент интересный. Сайт Kaggle, он, это сайт, который имеет дополнительную мотивацию. У UCI просто ты вот либо решил задачу, либо не решил там, и всё такое. А вот на Kaggle, там, если решил и лучше других, там конкурсы проводятся, конкурсы. Вот. То есть действующие конкурсы, есть прошедшие конкурсы, есть песочница, где есть наборы данных без конкурса. Ну можете взять и оттуда, никто это не зазорно. Тем более, может быть, первый раз вы это делаете. То есть вполне не зазорно это сделать. Так вот, что такое конкурсы, ребята? Значит, там вы делаете приложение, потом используете тестовую выборку для того, чтобы сделать прогноз на основе этого приложения или или идентификацию. Посылаете её на сайт UCI, там причём можно каждый день делать по пять отправок можно делать. Вот. И если вы занимаете первое место, вам даётся половина призового фонда, а он может быть до 100.000 долларов. А если второе место, тогда 50.000 делится на 1/3 и 2/3. Вы получаете 2/3 от оставшейся половины, а 1/3 получает тот, кто занял третье место. Ну, примерно такая вот система. Короче говоря, там есть монстры, ребята, на этом Кагле. Ну я не знаю, как их назвать, ну очень такие ушлые специалисты, которые вот действительно хорошо в этом разбираются, имеют очень хороший инструментарий, позволяющий решать эти задачи. На Кагле иногда встречаются задачи, которые, ну, в Excel, например, форма не помещается. Вот скажем, в Экселе миллион строк, а там форма 4 млн строк. Значит, я вводил в специальных случаях в систему Eidos, то есть делал интерфейсы CSV, стандарт файла, вводил. И я вам могу сказать, что где-то до 4 млн, 5 млн обучающей выборки система Eidos обрабатывает. То есть у неё её ресурсов, вот её реализации программной хватает для обработки таких массивов данных. А очень многие системы, типа SPSS, статистика, они на порядок меньше имеют возможности просто, даже на два порядка меньше, понимаете? Вот в системе Eidos я, например, недавно обрабатывал задачу, где 200 шкал. А в системе Статистика только 70 шкал самое большее, понимаете? Короче говоря, я мне удавалось решать задачи, которые другие разработчики, они просто не имеют инструментария для решения таких задач, у них нет систем программных, которые позволяют такие задачи обрабатывать, понимаете, такие объёмы данных. Система Eidos это позволяет.

Значит, я пробовал участвовать в этих конкурсах, мне интересно было, какая моя позиция там, где я там болтахаюсь между этими лидерами и теми, кто вообще там в концах там где-то. Значит, я могу вам сказать, что я где-то в серединочке там болтахаюсь, даже в верхней части. Значит, вот сейчас мы посмотрим. Вот Kaggle. Вот он я. Вот. У меня, я несколько раз участвовал в конкурсах. И самым высшим достижением... Сейчас я не помню уже. Надо войти туда, наверное, чтобы увидеть это.

Вот, соревнования. Вот я участвовал в соревнованиях. И вот, ребята, результаты. Смотрите. Значит, я в двух соревнованиях участвовал. В одном я занял место такое, позицию выше 85% участников. А в другом соревновании занял позицию выше 91% участников. Ну, это, как сказать вам, это, конечно, не лидер, до лидера там далеко. Вот, но это как бы неплохо, можно сказать так. Ну, не хуже других, в общем, вот так примерно. Значит, что я могу сказать? Что я не имею особых возможностей участвовать в таких соревнованиях, потому что, чтобы в них участвовать, нужно этим заниматься. Ну то есть берёшь, делаешь модель, а потом её совершенствуешь, понимаете? Улучшаешь её характеристики этой модели. А это можно делать очень большим числом способов. Можно делать разное число интервалов, можно делать адаптивные интервалы и равные интервалы. Можно там, ну, в общем, какие-то создавать шкалы, ну, скажем так, составные. То есть те шкалы, которые там имеются в исходных данных, их можно преобразовать во вторичные данные, и эти вторичные данные как промежуточный слой использовать. И тогда достоверность решения задач возрастает. В системе Eidos есть соответствующие возможности, прямо они все режимы, которые для этого предназначены, которые формируют сочетание факторов первого слоя, делают его классами, потом, значит, формируют сети на основе этих классов, как уже рецепторов, и формируют классы более высокого уровня, которые являются выходными для задачи, которая там ставится, и так далее. В общем, там есть, короче. Но это надо возюкаться, понимаете? У меня никакой возможности возюкаться нет, потому что у меня бешеная совершенно нагрузка. Огромная нагрузка. То есть я вот так вот сижу целыми днями, веду занятия. Значит, я сейчас вам покажу просто таблицу расписания, которое мне прислали, так вот. Ну вот. И когда мне тут что-то тут ещё делать? Я дай бог, чтобы вообще очухался вот после занятия и провёл следующее, понимаете? Потому что большая физическая нагрузка. Весь день я веду занятия. Вот. И, если честно, то не только в нашем вузе. А ещё в двух вузах. Вот. Ну это, конечно, моя проблема, но вот ещё один вуз, видите, Пермь. А вот ещё один вуз - это КГУ. Там довольно много дисциплин, между прочим. Это вот кафедра интеллектуальных систем. А это кафедра информационных образовательных технологий. То есть я профессор на этих кафедрах ещё. Вот. Как вот у нас есть страничка в университете. Вот. Так и у них есть страничка. Это вот страничка, значит, у них в КГУ, видите? И вот там веду ряд дисциплин. Вот моя почта, Skype, проекты. Всё тут есть. Вот. И есть наша страничка, видите, Аграрного университета. Вот. Дисциплины, которые я веду здесь. Ну я вам скажу так, что мало не покажется никому. Поэтому мне вот не до этого.

**4.2. Использование шаблонов и публикация**

Но у меня большое достижение было, знаете какое? Не на Кагле, а, значит, э-э, в чемпионате РАИФ 2017 я принял участие вот в этом чемпионате. И там было три номинации. Я в двух номинациях занял первое место, а в одной - третье, по мнению экспертного сообщества. И у меня первый рейтинг в России в области искусственного интеллекта и управления. Ну это, пожалуй, вот и являются достижениями. Вот. Ну, правда, сам этот критерий, он такой сомнительный, я бы сказал. Вот. Сейчас я покажу вам первый рейтинг. Вот, смотрите, рейтинг учёных России в области кибернетики, где относятся интеллектуальные технологии, теория управления. У меня первый рейтинг в этом по этой направлению науки. Просто вот в России, понимаете? По таким вот показателям многокритическим. Ну то есть что-то как-то где-то.

Значит, вам надо зарегистрироваться на Кагле и посмотреть, какие там есть наборы данных. Вот. И попробовать какой-то набор данных, насколько он удобен для ввода в систему Eidos, посмотреть просто. И попробовать решить задачу. Кстати, задача может быть связана с обработкой текстов, с обработкой табличных данных и числовых, и текстовых, и с обработкой изображений. Требования, чтобы эта задача ещё не решалась у нас, чтобы или хотя бы не решалась в таком же виде на этом же наборе данных, вот так. Вот по вину я, например, писал статьи. Но возможно, я писал на других наборах данных, не на этом наборе данных. Ну тогда можно исследовать. Я исследовал, как связано... О! Это я не решал. 130.000 отзывов. Ребят, вот такие объёмы информации их вполне можно обрабатывать в системе Eidos. Вы не пугайтесь. Я обрабатывал до 700.000 объектов обучающей выборки. То есть, ну, ну, может быть, конечно, некоторое время там она пошерстит, там поработает. Вот. Ну, конечно, надо CSV формат использовать. Вот. А по вину у меня есть статьи по вину. Сейчас я покажу на этом конец занятия. Короче, вам надо сейчас искать тему.

**5. Заключение**

На этом у нас занятие заканчивается. Какие у вас возникают вопросы, пожалуйста, можете задать. В общем, сейчас у нас задача, у вас задача выполнять эту вот инструкцию по разработке собственного приложения. Сейчас мы находимся на этапе четвёртый пункт выполняем. То есть вы смотрите, какие наборы исходных данных есть, ищете то, что вам было бы интереснее. И можете также сами создать набор данных или найти в другом месте где-нибудь, не там, где вот я рекомендую, а где угодно можете найти. Значит, у меня на сайте есть информация об этом, где есть наборы данных. Сейчас я вам покажу ссылочку, и вы потом можете это использовать. Вот, смотрите, вот страничка есть, и здесь тоже есть наборы данных. 50 лучших баз данных для машинного обучения тут и так далее. Я вам даю ссылочку на эту страничку. В общем, ищите, ищите, короче говоря, наборы данных и решаете, пробуйте сделать модель, оцениваете её достоверность. Не получается, приходим вот сюда на занятие, и на занятии занимаемся именно вот этим, ребята, именно этим. То есть мы смотрим, вы присылаете мне на почту, может быть, свои наборы данных, мы делаем модели, смотрим, я через облако, как угодно. И вы, если получается так разумно, что-то интересное, ну тогда вы делаете описание по шаблону. И потом дальше регистрируетесь в ResearchGate, размещаете и в РИНЦе. А сейчас можете начать эту регистрацию. Там есть описание, как это делается. Вот здесь вот у нас есть файлик, где знаю, вордовский файл, вот здесь описано, как эта регистрация осуществляется. И в Кагле осуществляете регистрацию для того, чтобы скачать данные.

Всё, самого хорошего, ребят, до свидания.
До свидания.
До свидания. В результате, в результате вы всё это освоите и будете уметь это делать. До свидания. До свидания.