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

***Российская Федерация***

**184 Лабораторная работа №8. По дисциплине Интеллектуальные информационные системы и технологии. 2020-11-06**

Eidos Lab 8: Project Theme Selection, Data Sources (UCI/Kaggle), and Publication via РИНЦ/ResearchGate

Резюме

I. Введение и детали занятия

Занятие представляет собой лабораторную работу №8 по дисциплине "Интеллектуальные информационные системы и технологии" для группы ИТ 18-23, проводимую 6 ноября 2020 года профессорами Луценко Е.В. и Аршиновым Г.А.

II. Посещаемость и коммуникация

Отмечается низкая явка студентов. Профессор Луценко подчеркивает необходимость обсуждать темы проектов во время занятий из-за нехватки времени на индивидуальную переписку, связанной с текущей аттестацией и оценкой.

III. Лабораторная работа №8: Выбор и обсуждение тем проектов

Студентам предстоит выбрать тему для разработки собственного интеллектуального облачного приложения в системе Eidos. Темы могут быть предложены студентами или выбраны из доступных источников (например, Kaggle), при условии соблюдения законодательства и этических норм. Допускаются темы, связанные с играми, при условии зрелого подхода к анализу. Профессора готовы обсудить выбранные темы во время занятия.

IV. Задание: Разработка и документирование приложения Eidos

Основная задача – разработать приложение в системе Eidos и описать его согласно шаблону. Необходимо установить и, возможно, обновить систему Eidos (инструкции даны). Задание направлено на оценку знаний, умений и навыков студентов, демонстрируя их понимание уровней данных, информации и знаний.

V. Публикация результатов исследования

Важной частью задания является публикация результатов исследования. Это необходимо как для оценки студентов (влияет на стипендию, входит в портфолио для ВКР), так и для отчетности преподавателей. Публикация осуществляется через ResearchGate (требуется DOI и корпоративный email для регистрации) и РИНЦ (Российский индекс научного цитирования). Для РИНЦ необходима регистрация в Science Index и получение SPIN-кода, а также заключение договора с РИНЦ. DOI можно получить, разместив работу как препринт или технический отчет на ResearchGate. Возможна публикация в соавторстве со старостой (у которого есть корпоративный email) или преподавателями.

VI. Поиск данных для проекта

Рекомендуется выбирать наборы данных (datasets) размером до 10 МБ из-за ограничений бесплатного хостинга ByetHost, используемого для облачных приложений Eidos. Основные источники данных – репозиторий UCI и платформа Kaggle, предлагающая как наборы данных, так и соревнования по машинному обучению с призовыми фондами.

VII. Заключение и дальнейшие шаги

Выполнение задания (разработка приложения, его описание и публикация) является обязательным для получения высокой оценки. Студентам рекомендуется активно приступать к работе, регистрироваться на необходимых платформах (РИНЦ, ResearchGate) и задавать вопросы преподавателям.

Детальная расшифровка текста

I. Введение и детали занятия

Доброе утро, Георгий Александрович.

Здравствуйте, Евгений Вениаминович.

Доброе утро, ребята и девушки.

Здравствуйте.

Доброе утро.

Ну, сегодня у нас уже 6 ноября 2020 года. И сейчас первая пара, 8:00–9:30. Лабораторная работа номер 8 с группой ИТ 18-23 по дисциплине "Интеллектуальные информационные системы и технологии".

Занятия ведут профессор Луценко Евгений Вениаминович и профессор Аршинов Георгий Александрович.

Вот. Такие вот дела.

II. Посещаемость и коммуникация

И у нас по плану сегодня, ребята. Вы ж мне многие писали письма, присылали свои темы, которые вы выбрали. И я некоторым писал, что на занятии будем рассматривать, на занятии, потому что совершенно нет времени. Мы вот сейчас с Георгием Александровичем сидим там ночами, заполняем всякие аттестации, оценки. Да, Георгий Александрович?

Вот. Так что особо так не разгонишься, чтобы ещё дома там смотреть ваши задачи. Ну сейчас мы готовы. Сейчас занятие. Мы готовы это сделать. То есть, если у вас у кого-то есть вопросы, то, пожалуйста, по вашим работам, по вашим темам.

Значит, во-первых, давайте так. Кто тут у нас староста-то, господи? 17... Чебенко, да?

Да, я.

Вот здесь, Дима. Значит, ну тогда скажи мне, почему так мало людей у нас? В два раза меньше, чем в группе.

Не знаю, я напишу сейчас.

А кто знает? Ты должен знать это. Ты должен это знать.

Наверное, спят.

Ну вот. Значит, значит, ты уже знаешь. Получается, что ты уже знаешь. Ну я тоже так подумал, что, скорее всего. Но надо просыпаться уже, уже занятия.

Вот. И, значит, это первое. Сейчас может и прийти проверка, бывает, иногда заглядывают. Причем они иногда заглядывают на секунду, там даже доли секунды. Вот раз, заглянули, что я говорю, вы сидите, они раз и сразу исчезают. Но я замечаю, что там мелькнуло.

III. Лабораторная работа №8: Выбор и обсуждение тем проектов

A. Процесс выбора темы

Так, теперь давайте обратную связь в чате мне. Кто делает своё приложение? Плюсик поставьте. Кто не делает – минус.

Или скажите просто, я вот по очереди, я не знаю, как вам, как здесь это организовать. Ну могу спрашивать каждого, если хотите.

А вот можно спросить насчёт приложения?

Конечно, конечно. Мы же для этого сейчас и сидим здесь.

А вот там тему нужно свою придумать или выбрать среди уже готовых?

Там можешь и свою, а можешь и выбрать.

Угу. Хорошо.

Ну, выбрать-то из тех, которые есть в Kaggle или у ЦИА, или на каких-то сайтах с данными.

B. Ограничения по темам и примеры (Игры)

Там суть-то в чём, ребят? Эти темы, они могут быть самые различные, но они не должны противоречить законодательству и морально-этическим нормам. Вот единственное ограничение. Не жёсткое, я бы сказал, ограничение.

Очень многие ребята, я вот вижу, они очень многие выбирают темы, связанные с играми. То есть у них ещё игровой период не прошёл. Продолжается. Ну я в этом ничего плохого не вижу. Даже скажу больше того, вот я вот сейчас только что переделывал систему и разместил новую версию на сайте. И вы думаете, что, мне это не интересно, что ли? Мне это интересно. То есть я бы сказал так, что, в принципе, я даже, ну, так скажем, вижу сходство с играми. То есть я вот сижу, что-то там делаю, у меня получается, мне это нравится. То есть, почему бы и нет?

Тем более игры, они довольно есть такие умные, и очень многие игры имитируют реальность. Неплохо. И я вот могу сказать, что меня поразил один студент, ну, несколько лет назад. Он сделал приложение по анализу каких-то ситуаций там в какой-то игре. Я прочитал, меня просто поразило, знаете что? Насколько зрелое понимание, восприятие у этого парня было. То есть я понял, что точно так же он может описать и ситуации в жизни, в фирме, в семье, там, где угодно, понимаете? Здорово, здорово, я бы сказал, вообще здорово просто. Мне понравилось, как он это сделал.

То есть я совершенно не против этого, пожалуйста. Я против того, чтобы из игр не вылазили или вылазили только тогда, когда уже прямо с голода падаешь или надо в туалет прямо срочно бежать. Конечно, это уже, ну, перебор. Должно быть разумно всё. В разумных формах. Как говорят, очень хорошее есть высказывание такое, что всё должно быть в меру. Должна соблюдаться мера во всём.

C. Текущая проверка статуса (отзывы студентов)

Так, вот Анастасия поставила плюсик. Одна такая у нас. Больше таких нет.

Я упустила момент, когда мы темы разбирали.

Да мы темы до сих пор разбираем. Это никогда не поздно. Можете сейчас тему выбрать. Мы её посмотрим.

Ну и, не знаю, как вас... Ну давайте, вот вы видите этот список?

IV. Задание: Разработка и документирование приложения Eidos

A. Установка и обновление Eidos

Значит, вот здесь вот у нас есть, смотрите, задание-инструкция для учащихся по разработке собственного интеллектуального облачного Eidos-приложения. Кликаем по ней и получаем текст. И этот текст я вам послал в чат. То есть не текст, а ссылочку послал в чат.

Что мы здесь видим на этом, в этом тексте? Что мы должны получить информацию о самой технологии интеллектуальной. Мы получили. Мы должны установить у себя систему Eidos. Вот это я не знаю, сделали вы или нет, ребята?

Значит, давайте сейчас в чате, в чате ставите, пишете: установил, установил, установил там. А если нет, то пишете: не установил, и какая причина? Какие проблемы?

Вот, даже уже и обновил. Ну вообще. Вот.

Ну я можно целиком не скачивать, а можно обновить. Я вам рассказывал, как обновить быстро, да? Можно очень быстро это сделать, легко. Вот здесь вот, смотрите, скачать и запустить систему Eidos. А здесь есть файл обновления. Вот его клацаете, кликаете здесь просто. Он скачивается. Потом берёте его, разворачиваете в папочке с системой с заменой всех файлов. Ну для того, чтобы все файлы заменились, естественно, система не должна быть запущена. Иначе она не заменится, исполняемый модуль системы.

Обновление – это что? Это обычно новый исполняемый модуль и какие-то ещё изменения. Вот сейчас, например, я внёс изменения в алгоритм принятия решений. Немножко там нашёл некоторую неточность, исправил её. И конвертер немножко улучшил CSV в DBF конвертер. И ещё в режиме 2.3.2.2 чётче сделал... Я вчера столкнулся с тем, что один студент задал адаптивные интервалы и специальную интерпретацию текстовых полей. И она у него не сработала, не выдала ему специальную интерпретацию, то есть слова не стала рассматривать как признаки. Я, значит, зашёл в саму систему, в исходный текст, посмотреть, что ж такое там, почему она не работает. А там, знаете, что написано, ребята? В исходном тексте прямо. А там написано... Я бог знает когда там, не знаю, когда, может, несколько лет назад даже. Я там написал себе. Я вот такие пометочки делаю себе, видите, вот? Вот это я себе пишу. Что надо обратить внимание на это. Значит, если у нас задан второй тип интервалов, то есть разные интервалы с равным числом наблюдений, то в этом случае не применять специальную интерпретацию текстовых полей. А у меня технически это было возможно. То есть вот так задашь этот второй вариант интервалов, и технически можно задать специальную интерпретацию, и там параметры задать и запустить. Она не работает в этом режиме. Сейчас я сделал так, что она и не должна работать, потому что она... это просто одно другому противоречит логически. Короче говоря, что я сделал? Я сейчас взял и изменил так этот вот интерфейс этого режима 2.3.2.2, что физически невозможно задать специальную интерпретацию текстовых полей, если задан вот этот параметр. Просто это вообще невозможно сделать. Даже и я не могу это сделать. То есть просто их там нету этого.

Вот смотрите, сейчас показываю. Это я сделал 15 минут назад. Ну, во-первых, вот показываю вам алгоритм. Я внёс в него изменения. Вот этот алгоритм. Вот здесь была стрелочка вот сюда. А я здесь написал выход. Почему? Потому что здесь вот у нас есть прогнозирование. Да, вот прогнозирование. Если всё нормально, тогда выход. А раньше был выход отсюда и опять на прогнозирование, понимаете? А второй раз прогнозировать какой смысл? Это имеет смысл только тогда, когда мы заменили какие-то там значения факторов на другие, тогда опять надо прогнозировать. А если мы не заменили, взяли вот, исключили, у нас всё получилось, ну просто вышли и всё. А здесь вот у меня было опять вот сюда выход. Ну это, конечно, незачем было. Вот я изменил алгоритм этот. Чуть-чуть уточнил.

А вот в этом режиме 2.3.2.2, смотрите, как сейчас работает. Вот видите, сейчас вы можете задать специальную интерпретацию текстовых полей. Видите, работает всё. Вот. А если я возьму, задам разные интервалы, всё, вообще исчезает эта строчка. Вот обратите внимание, раз. Видите, можно применить специальную интерпретацию текстовых полей. А если я задам разные интервалы, то вообще этой строчки нету. То есть я даже, в принципе, технически не могу задать специальную интерпретацию. Даже нет места, где поставить там птичку, понимаете? Вот что я сделал сейчас. Вот. И применить сценарный метод. Применить, чистенько всё работает. Применить, не применить. Вот. Применить сценарный метод, не применить. Всё чётко сделано. Вот это я сейчас только что сделал.

А почему я это сделал? Потому что, я же говорю, вчера студент задал эти параметры, ну, некорректные он задал параметры, конечно. Но он-то не знал, что они некорректные. Откуда он знает? И я забыл, что они некорректные. Когда вот у него не получилось, тогда я вспомнил. Ну, думаю, надо сделать так, чтобы вы не могли это задать такие параметры, взаимоисключающие.

B. Цель задания (Знания, Умения, Компетенции)

Вот, устанавливайте новую версию. Ну теперь начинаем, ребята, рассматривать вот это вот задание-инструкцию. Во-первых, почему она называется задание, ребята? Потому что это и есть задание. Вот сейчас, пожалуйста, послушайте просто меня внимательно.

Значит, что мы должны вам давать, преподаватели? Что мы должны вам давать? Знаете, нет? Что мы должны вам давать?

Знания.

Деньги должны давать. Ну, можно дать денег, можно дать удочку. Лучше удочку, да? Вот знания, конечно, это удочка. То есть, но я хотел вам сказать, что сразу вот я сейчас вспомнил смешной очень случай из своей практики. Когда-то, когда только пришёл в аграрный, это было очень давно, 80 лет назад, в университет, то, значит, там не было часов. Я как вклинился в коллектив, там трудно было меня на ставку найти часы. И мне давали разные дисциплины, самые различные: алгоритмы, структуры данных, там, информатику, там, вот такое, математические основы информационных технологий. И, короче, дали мне информатику. Я, значит, веду эту информатику, рассказываю, что компьютер состоит там из монитора, клавиатуры, там. Вот. И потом мы, значит, изучаем там Word, Excel, ну такое что-то. И, значит, и все сидят студенты первого курса и изучают, выполняют задания, всё нормально. Одна девушка сидит задом к компьютеру. Ну вообще задом, вот как сказать? Правда, ко мне лицом. То есть у неё она в таком месте сидела за компьютером, что если вот повернуться к компьютеру, то будет ко мне задом. Вот. Я её спрашиваю: "А вы почему не выполняете задания? Все же выполняют. Вот вам дано задание, давайте выполняйте". Вот её ответ, он настолько меня поразил, я вот запомнил его. И просто как, я даже не знал, что ей сказать. Вот я сейчас вам этот ответ воспроизведу. Она мне говорит: "А наш преподаватель информатики в школе не вложил в меня знания компьютера". Вам нравится это, нет?

Я говорю: "А-а, а, ну, если он не вложил, ну тогда что я могу сделать?" Я говорю: "Ну тогда давайте". Значит, потом стала объяснять, что вообще-то знания компьютера не вкладывают, понимаете, преподаватели. Они создают условия для того, чтобы вы сами это освоили. Ну это примерно как вот преподаватель не вложил умение плавать или не вложил умение ездить на велосипеде. Его невозможно вложить. Вы же не роботы, правильно? Если бы вы были роботы, вам можно было бы туда вставить там какую-нибудь флешку, перекачать там программу вождения велосипеда, и всё, и вы бы водили велосипед там прекрасно, на нём как циркачи ездили, допустим. Но это ж не так. А как? Чтобы научиться ездить на велосипеде, нужно сесть на него и ехать, понимаете? То же самое касается плавания. Чтобы научиться плавать, нужно прыгать в воду, а не на лавочке лежать и руками махать и ногами, понимаете?

Даже есть такой же анекдот аналогичный, что прыгают с вышки в бассейн, где воды нету. Спрашивают: "Вы что, вообще что ли? Это же можно поубиваться". Они говорят: "Ну нам сказали, что когда вот мы научимся, тогда нальют воды". В таком плане. Ну то есть это ж понятно, что для того, чтобы научиться работать на компьютере, нужно поворачиваться к нему лицом и сидеть и работать.

C. Уровни данных, информации и знаний

Так вот, ребята, мы должны вам давать знания, умения и навыки. А сейчас ещё добавляют и соответствующие необходимые компетенции, которые там в рабочих программах прописаны, в учебных рабочих планах. Вот. И где дают знания? Знаете? На лекциях. То есть когда вот мы вам что-то рассказываем, вы получаете данные, информацию и знания. Данные вы получаете, знаете когда, ребята? Слушайте. Вы мне спрашиваете: "Нет, не знаем" или "Знаем". Вот когда вы получаете данные, когда вам рассказывают? Кто-нибудь знает?

Ну ответьте мне, знает кто-нибудь или нет? Скажите: "Нет, не знаем" или "Да, знаем".

Ну, я предполагаю, что не знаем.

Ты знаешь, что ты не знаешь, да? Ну, короче говоря, данные, ребята, это... я вам вообще-то рассказывал соотношение этих понятий: данные, информация, знания. На лекциях должен был рассказывать. Значит, данные – это неосмысленная информация. То есть если вот мы вам рассказываем что-то, а вы не понимаете, ну, не въезжаете, как вот вы выражаетесь. Вот. Там лагаете там, не знаю, в общем, какие у нас там слова есть, когда вот вы не понимаете. Вот. Новенькие.

Вот тогда вы получаете данные. А как тогда сдавать экзамен? А запросто, взял, вызубрил, там, запомнил фразы. Обычно у девчонок это отлично получается. У ребят похуже. Значит, ну, в принципе, на экзамен приходят и даже меня это удивляет иногда, прямо вот чётко всё рассказывают. Спросишь что-нибудь, не дай бог, вот даже по этой же теме, ничего не могут сказать, понимаете? Это что означает? Вызубрили. Вызубрили, очень неплохо получается. Я даже так не умею, если честно. Ну, там было несколько раз, когда там английский язык сдавал там. Вот, тогда, ну, эти вот диалоги я выучивал, конечно, вызубривал. Вот. Ну вообще-то, это мне против шерсти, мне это очень не нравится это занятие – зубрить.

Вот. А информационный уровень обучения – это обучение, когда вы понимаете, о чём я говорю. Вот сейчас вот вы понимаете, о чём я говорю, нет? Ребят? Ну?

Это вы так думаете, что вы понимаете. А там мы разберёмся, поняли вы или нет. Вот. Так вот, ну это шутка, конечно. В общем, вы, конечно, понимаете, о чём я говорю. Это означает, что вы получаете информацию. Потому что информация – это осмысленные данные. То есть данные, которые вы поняли уже содержание этих данных. То есть это более высокий уровень обучения. То есть если вы приходите на экзамен и понимаете вопросы, то вы можете и ответить, не вспоминая, что там написано было или что я говорил там, не вызубривая. Вы просто можете своё понимание изложить, и это будет нормально, это будет отличная оценка.

Вот. А знания как получаются, ребята? Знания получаются только тогда, когда вот эта вот информация, которую вы понимаете, осмысливаете, она для вас важна по каким-то причинам. Значит, по каким причинам она может быть для вас важна? Если вы используете её для достижения своих целей каких-то, которые вы ставите. Если вам это интересно, нравится. В любом случае, знание – это не просто логические какие-то структуры, которые вот вы там понимание, связанное вот с информационным уровнем. А это значимая информация для вас. То есть это то, что для вас важно. Вот если я вам рассказываю что-то, а вам это не важно, то вы не получите знаний, вы будете учиться на информационном уровне. То есть вы будете так всё понимать более-менее, но знаний не получите. Потому что знание – это не просто информация, а это информация, которую вы используете для достижения целей. Вот. Которую вы активно, так сказать, её пытаетесь приобрести и использовать. То есть знание – это такое креативная такая штука. То есть это можно сказать так, креативная информация. Информация, которую можно использовать.

Вот. Я спрашиваю вас: а у вас, вам это интересно, что я рассказываю? Ребят? Ну? Вот сейчас вам интересно?

Очень познавательно.

Вот. Значит, я вам сейчас очень интересную вещь скажу. Кому интересно, тому один балл на экзамене. А теперь скажите, интересно вам или нет?

Интересно.

Очень даже интересно. А если сказать, что очень интересно, тогда что, два балла получаешь что ли на экзамене?

Интересно.

Вот тут начинается уже... Очень даже интересно, да? А что, а если я скажу, что вообще просто обалдеть, как интересно. Ну вообще вот, интереснее, чем на всех других занятиях. То тогда получаешь три балла плюс к оценке на экзамене и ещё 1000 рублей.

Ну это я так подшучиваю. Тогда становится вообще очень интересно. А если 1000 долларов, тогда ещё интереснее становится. Да, ребята?

Наверное.

Вот. Так вот, если вас заинтересовать, то тогда начинаете вгрызаться в эту в эту информацию, начинаете прямо вот её пытаетесь её получить, там, пытаетесь её освоить. Потому что вам становится это важно для вас. Вот это уже тогда вы приобретаете. Вот при таком состоянии, когда вы не просто так вот сидите там и слушаете, что там по телевизору рассказывают там, на экране там или на телефончике. А когда вы начинаете прямо вот увлекаться этим, когда вам это становится очень интересным, важным для вас. Вот. А как преподаватель может сделать так, чтобы вам это было важно для вас?

Вот у преподавателя есть разные методы для этого. Один метод какой? Знаете?

Поощрение.

Ну, метод кнута и пряника, да. То есть это старинный метод такой, довольно эффективный. Значит, люди вообще-то, они в этом смысле, можно сказать, поддаются дрессировке. Ну как вот можно собачке дать там что-то скушать и похвалить. А не сделал, значит, можно её хлопнуть там, шлёпнуть. И она потихоньку начинает понимать, что надо всё-таки вот делать то, что говорят. Примерно так же и люди. То есть такой уровень, конечно, тоже возможен мотивации.

V. Публикация результатов исследования

A. Важность публикации (для студентов и преподавателей)

Ну, значит, преподаватель есть такой такие варианты. Он может сказать так: "Ребята, вот не будете слушать, что я рассказываю, вам, так сказать, экзамен вы не сдадите, понимаете? Вы не сможете решать задачи".

Вот как я сейчас вот я преподаю вам, вот так я определю, освоили вы знания или нет? Знания, умения и навыки. Да никак не определишь. Вот сейчас вот, когда была у нас в аудиториях занятия, я видел, что вы сидите за компьютером и работаете. А сейчас я не вижу. Я вижу там буковки вот на экране: Шевцов там Дмитрий, там ИТ 1823, там Чебенко Дмитрий. Вот, Сергей Генералов. Вот я вижу ваши фамилии просто и имена, и всё. Откуда я знаю, что вы там освоили, не освоили?

Значит, мне нужно проверить это каким-то образом. Вот поэтому мы вам даём задание, ребята, вот это. Это задание, оно предназначено для того, чтобы проверить, как вы освоили знания, умения и навыки, которые вам преподавали. Знания мы преподавали на лекциях. Вот. А умения и навыки на лабораторных работах. То есть я рассказывал вам, показывал, как решаются задачи различные. Довольно подробно всё объяснял. И последнее у нас было занятие, где я объяснял, как исследуются изображения, спектральный анализ изображений. Интеллектуальный причём спектральный анализ. Было дело, да?

Вот. Так вот, это мотивация. То есть я вам говорю, чтобы получить положительную оценку, вам нужно посещать занятия. А вас здесь 2/3 только сейчас, 1/3 нет вообще нету. Поэтому они, конечно, не знают, о чём я сейчас рассказываю, абсолютно не подозревают. Я вам говорю: "Вот, ребята, слушайте внимательно. Это не шутки. То, что я сейчас вам рассказываю, это я совершенно серьёзно говорю. Значит, хотя я с вами там так в вольной форме беседую, иногда там подшучиваю, но сейчас я совершенно серьёзно говорю. Это задание. Задание. Вам надо его выполнить это задание. Иначе вы получите три балла, понимаете?" Вот.

Как оценивается это задание? Если вы выполните это задание полностью, то получаете пятёрку. Если не полностью – четвёрку. Если выполните совсем-совсем маленькую часть, тогда удовлетворительно. Вот таким образом оно оценивается.

Что это за задание? Оно состоит в том, что вам нужно разработать своё приложение в системе Eidos. Это как минимум. Второе. Нужно описать это приложение в соответствии с шаблоном. То есть используется шаблон описания. То есть это типовой проект, так сказать, типовой способ изложения хода и результатов исследования, проведённого в системе Eidos. Здесь несколько примеров приведено из разных областей. Вот. И после того, как вы это всё опишете, вы должны разместить это... Ну, это делаю я конкретно, вы в Eidos-облаке, то есть на FTP-сервере системы Eidos. Это делается непосредственно в самой системе Eidos.

B. Платформы для публикации (РИНЦ, ResearchGate)

А также очень желательно разместить ResearchGate и РИНЦе.

C. Требования к регистрации (РИНЦ, Science Index, DOI)

Вот. Соответственно, вам надо там зарегистрироваться. Я, значит, просто вам настоятельно рекомендую этим заняться незамедлительно. Ну то есть прямо вот, когда у вас появится там свободное время после занятий, сразу начинать регистрироваться в РИНЦе. Что такое РИНЦ, ребята, знаете, нет?

Нет.

Вот вам везёт, кошмар. Ничего не знают. Много знаний – много печали. Это кто сказал, знаете, нет? Кто сказал царь Соломон, мудрейший из всех людей, живущих на Земле.

Значит, вот, ребята, это вот РИНЦ. И он же и eLibrary. Значит, ну, eLibrary – это понятие несколько более, значит, более широкое, чем РИНЦ. eLibrary – это российская библиографическая база данных, самая большая, которая есть в России. Там огромное количество статей, журналов. Ну вот, скажем, всего журналов 70.000, российских – 17.000, выходящих прямо сейчас – 14.000. Вот. А вот в РИНЦ индексируется из них 6.000, видите, 5.900. Что это значит? Что РИНЦ – это подмножество eLibrary, ну, скажем так, более престижное. То есть те журналы, которые более серьёзные, которые более котируются, более высоко, они вот помещаются в РИНЦ. А остальные вот, причём их 2/3 примерно, они не входят в РИНЦ, но в eLibrary входят.

И здесь очень много статей. В открытом доступе журналов 7.000, российских – 7.000. Вот. А теперь смотрите, значит, общее число публикаций, я даже не могу понять, сколько здесь, 35 млн публикаций, ребята. Вы понимаете? 35 млн публикаций. Из них с полным текстом 13 млн статей с полным текстом. А вот в открытом доступе уже поменьше – 6 млн статей в открытом доступе. Что такое открытый доступ? Это значит, что эту статью можно прочитать бесплатно. То есть можно её найти и прочитать здесь вот в РИНЦ. У него есть свой аппарат поисковый.

D. Процесс публикации и соавторство

Значит, ну а теперь, значит, смотрим дальше. Значит, ну, в общем, это огромная, ребята, база данных, в которой находится огромное количество журналов и статей. Самый большой в России, самая большая в России библиографическая база данных. Вот здесь вам надо зарегистрироваться. И также надо зарегистрироваться в системе Science Index. Зачем это нужно? Значит, слушайте внимательно. Вот вы выполняете это задание. Значит, в чём состоит это задание? Оно состоит в том, что вы проводите небольшое исследование, такое учебное, на данных, которых вообще, э-э, сколько угодно, скажем так, этих данных, огромное количество. Это сейчас мы потом начнём рассматривать, где эти данные брать. Данные для машинного обучения. Их огромное количество уже прямо в базах данных для этой цели созданных. А и ещё просто в интернете полным-полно данных, которые можно использовать для этих же целей.

Вот. И, значит, потом описываете это исследование. Описываете, как оно проводилось, какие получились результаты. А потом публикуете. Публикуете. И вот чтобы опубликовать это исследование, вам надо зарегистрироваться в ResearchGate и в РИНЦ.

Значит, как осуществляется публикация? Есть небольшая инструкция у меня по этому поводу. Сейчас я вам эту инструкцию ссылочку дам на неё. Вот вам ссылочка.

Возникает вопрос такой: а зачем оно мне надо вот? Ну, во-первых, потому что я вас мотивирую, и вы не получите хорошие оценки, если это не будете выполнять. Во-вторых, потому что это очень интересно, даже некоторые говорят, что это вообще безумно интересно. Вот. Но самое главное, это всё так, между прочим. А самое главное, если серьёзно говорить, оно вам просто нужно. Вот это вам действительно нужно. А для чего вам это нужно? Вы хотите получать стипендию? Хотим, да, говорите? То есть я уже о деньгах заговорил. Если вы хотите получать стипендию, вам нужно иметь какие-то достижения. Вот публикация статьи является таким признаком того, что студент старается, учится, и должен быть поощрён тем, что у него будет стипендия.

Вот. Потом, когда вы будете заканчивать обучение, это на самом деле не так уж долго ждать до этого времени, как вам кажется. У нас будет защита дипломной работы, ВКР. И там вы должны представить портфолио. Что такое портфолио, вы знаете, ребята, нет?

Список достижений. Знаем.

Да, да. И вот в портфолио одними из самых главных ваших достижений, которые в портфолио указываются, являются публикации. Ну, дипломы, разные аттестаты, э-э, там, если вы выступали на каких-то конференциях там или занимали какие-то призовые места на каких-то олимпиадах там или на конференциях. Вот. Студенческих.

И вот, значит, вам нужны публикации, короче говоря. Они нужны не только вам, они нужны и преподавателям. Преподаватели периодически проходят конкурсы. И на конкурсе смотрят, а сколько ж у него публикаций за 5 лет? И у него нету. Тогда говорят: "Да он чем он вообще занимался эти 5 лет? Профуфукал вообще их, так сказать. Ну он, может быть, вёл занятия, но наукой не занимался". А преподаватель должен заниматься и обучением, и преподаванием, и обучением, преподаванием, и наукой должен заниматься. Это по должностным обязанностям, понимаете? Вот. То есть получается, что преподаватель даже не может пройти по конкурсу, если не будет у него публикаций, не может должность занять доцента, к примеру, или профессора.

Члены совета диссертационного тоже, если у них не будет достаточно публикаций, они тогда исключаются из совета. Я видел, как это происходит. Ну обычно это происходит, когда профессор становится совсем старенький, больной, и у него изменяется отношение к жизни. И перестаёт его интересовать этот совет эти диссертационные. Становится ему всё это, ну как бы, равнодушным он к этому становится, вообще к этой деятельности научной. И он практически перестаёт работать. И вот когда он перестаёт работать, то это выражается в частности в том, что у него исчезают публикации, нет у него публикаций. Тогда ему говорят: "Ну раз ты уже фактически не являешься учёным, когда-то ты был известным учёным, у тебя было много публикаций, много книги были, статьи там, патенты". Вот уже прошло 5 лет, а ты ни одной не написал книги, ни одной статьи. Во что ты превратился? Ты посмотри на себя вообще, чем, кем ты стал. Вот. И, значит, ну так не говорят, конечно. Но смысл такой примерно в этом. И говорят: "Всё, ты не соответствуешь критериям, которым должны удовлетворять члены учёного совета, диссертационного совета". И поэтому мы тебя исключаем. Ты вообще на заседании совета ты спишь там, например. Ну и всё, понимаете? То есть вот на этом заканчивается.

То есть это показатель, то есть публикация, ваши публикации – это показатель того, что вы учились не зря, что вы действительно приобрели знания, умения и навыки. А для нас с Георгием Александровичем самое главное в этом что? Если вы сможете вот этот пройти этот путь с нашей помощью, я подчёркиваю, на всех этапах: на этапе выбора темы, на этапе описания, на этапе размещения, везде мы вам поможем, везде, на всех этапах. Вы пройдёте полный цикл научного пути, то есть пути исследования: от постановки исследования до его выполнения этого исследования, реализации, и описания, и публикации. То есть это тот, то, что делают вообще-то учёные. Они проводят исследования, описывают их и публикуют. Вы пройдёте полный цикл этого, понимаете? Вы приобретёте соответствующие знания, умения и навыки, которые вам в будущем очень пригодятся. Особенно, если вы будете заниматься наукой. Но если даже не будете, то у вас это входит в вашу программу обучения. Вы получите необходимые компетенции.

VI. Поиск данных для проекта

A. Рекомендации по размеру данных

Значит, э-э, не берите... Я вам убедительно рекомендую, просто вот убедительно. Не надо выбирать данные, которые больше 10 МБ исходные данные. Потому что вы замучаетесь обрабатывать их. Это может занять у вас сутки, например, понимаете? То есть это не требуется абсолютно в данном случае. Потом, если будет интересно, можно это сделать. Но это когда будет возможность, время там и так далее. И ещё одна причина есть. Вот этот FTP-сервер ByetHost, э-э, ByetHost, значит, американский, он изумительный сервер. Значит, там неограниченный объём памяти, без рекламы, высокая скорость, 1000 Мбит/с. Все SQL там, PHP, всё там, Java, всё там есть. Э-э, значит, э-э, ну, в общем, скоростной, без рекламы, неограниченное пространство. То есть вообще просто сказка. Вот. Однако у него тоже, вот на солнце тоже есть пятна. У него есть вот такое вот ограничение, что если туда запишешь файл больше 10 МБ, он сразу его стирает, молча просто. Да, и ещё очень интересный момент. Этот хостинг, ByetHost, он может, э-э, то есть там нет требования, чтобы на все файлы были ссылки. То есть это не только для сайтов он годится, но и для передачи данных, хранения данных он годится. Вот, просто как вот облако. Почему? Потому что можно разместить там файлы, на которых нет ссылок, понимаете? И он к этому спокойно относится. А вот на Яндекс, например, э-э, там так не сделаешь уже. Я имею в виду не облако Яндекс, а сайт Яндекс. Если вы сайт делаете на Яндексе, то он будет стирать просто файлы, если на них нет ссылок.

B. Источники данных (UCI, Kaggle)

Раньше он был хороший, потом его купил Яндекс, он стал такой вот. Я перестал пользоваться им. Лучше всего, чтобы там килобайтами измерялось, не мегабайтах, а там килобайтах. Ну, 1-2 МБ максимум. Больше не нужно, я вам просто вот советую, рекомендую.

Смотрим UCI портал. Самый древний набор данных – девяносто пятый год, ребята. Этот портал древний, понимаете? То есть он давно существует. И вот вы можете посмотреть, что здесь у нас есть. Он постоянно пополняется. И вот здесь мы можем увидеть, что у нас в 2020 году тоже существенное пополнение. Вот девятнадцатый год и так далее.

Теперь давайте сейчас зайдём на мой сайт, во второй раздел, кликнем на ссылочки "Скачать и запустить систему Eidos" и на Ctrl+F наберём и наберём UCI. Вы видите семь, ребята, статей. Видите? Где в названии статьи есть UCI.

Теперь смотрите, ребята, вот сайт Kaggle. Сайт Kaggle. Это сайт, э-э, создала фирма, э-э, самостоятельная с таким же названием. И он очень быстро приобрёл, э-э, огромную популярность, гораздо выше, чем у UCI сайта. Чем это обусловлено? Что они придумали такого, чтобы так вот их такая у них бешеная популярность, ребята? Они придумали простую вещь – мотивацию создали. Кроме... На UCI там просто датасеты, да? А на Kaggle – конкурсы, ребята, конкурсы, в которых есть призовые фонды.

Вот смотрите. Ребят, вам видно, нет, вот это? Что здесь написано? Какой призовой фонд? Видите?

100.000?

Да. Да. 100.000 долларов, ребята. Лидеры. Пять лидеров получат эти 100.000, разделят между собой, но не сами, а там есть процедура. Ну, те, кто первое место займут, они получат больше, те, кто последнее – меньше, но тоже там тысяч три-четыре, пять тысяч они получат долларов. Те, кто пятое место занял. Остальные ничего не получат.

Есть на... Значит, ну я могу вам сказать, что это довольно сложно попасть в лидеры, потому что там есть ребята, которых я не знаю даже, как их назвать вообще, ну, зубры такие вот. У них, видите, здесь вот такое вот есть вокруг колечко такое. Это колечко, чем оно сильнее замкнуто, тем, значит, этот уровень выше у этого участника. А здесь вот даже таких-то два колечка есть, я вижу. Вот. Да, кстати, здесь есть и много из России. Вот я сейчас только Афанасьев Дмитрий, видите, вот? Из России. Вот. Хотя, может быть, не из России. Вот, пожалуйста, мастер соревнований, видите? Токио. Вот. Мастер соревнований. А бывает иногда гроссмейстеры здесь участвуют. Это великий мастер соревнований. Мастер Kaggle. Вот. Я видел гроссмейстера как-то раз. Вот, здесь вот на Kaggle.

То есть ведутся рейтинги. Ну я тут не участвую в этих соревнованиях, могу сказать. Значит, я пробовал немножко в некоторых, несколько там, два раза. Ну, получаются слабенькие результаты, и я, в общем, не стал этим заниматься. Почему? Потому что, чтобы получить хороший результат, нужно стараться, ну, трудиться надо, время тратить надо. Вот. Ну я могу вам сказать так, что вот в этих рейтингах я там куда-то, то там в первой трети, то в последней трети, то есть вот такое вот, в общем, не особо. Вот. А в России я участвовал в чемпионате, открытом чемпионате России, я занял два первых места из трёх. А одно занял третье место в одной категории, номинации. Но там, правда, по оценкам независимого экспертного сообщества. А первое место дали тому, кто получил пятнадцатый рейтинг по оценке экспертного сообщества. Ну то есть там Сколково, это, в общем, там, тут не пробьёшься туда. Ну по оценкам сообщества, ну как бы, знаете, вот приз зрительских симпатий. Лазарев получил первое место, а этот самый, Кончита Вурст получила первое место по этому, по решению судей. А Лазарев по зрительским симпатиям, а по решению судей второе место он получил. Знаете, да, эту историю?

Так вот Kaggle создал мотивацию тем, что платят эти вот призовые фонды. То есть это создало такой азарт определённый во всём этом. Во-вторых, э-э, э-э, на UCI вы просто данные скачиваем и можем сами посмотреть там у себя дома, что там получается. А на Kaggle мы можем эти данные отправить на Kaggle результаты моделирования. И они нам определяют рейтинг наш, понимаете, по этим результатам, насколько у нас хорошо получается. То есть мы видим дополнительные функции на Kaggle, которых у UCI нету. Во-первых, связанные с мотивацией, во-вторых, с определением вашей позиции в мире среди специалистов в этой области, искусственного интеллекта. И третье, значит, там есть конкурсы уже прошедшие, есть конкурсы текущие, и есть, э-э, данные, размещаемые вне конкурсов. Вот. Значит, а просто, значит, потренироваться, чтобы люди могли просто потренироваться или как-то там позаниматься этой тематикой.

VII. Заключение и дальнейшие шаги

Значит, я вам советую, ребята, искать на Kaggle, на UCI исходные данные, соответствующие вот этим требованиям нашим, которые я уже вам сказал, какие, чтобы этих задач ещё не было решено вот здесь, смотрим, чтобы они были не очень большие и не противоречили морально-этическим нормам и действующему законодательству. Вот, ребята. И давайте, не откладывая, это делаете. Находите какие-то данные, скачиваете.

Так, ребята, у нас конец занятия, насколько я понимаю, да? Правильно?

Да.

Всего самого хорошего вам. Если будут вопросы, можете мне писать на почту. Но я вам, скорее всего, отвечу, что это на занятиях мы будем рассматривать. Почему? Потому что очень большая перегрузка. Вот. А вот появилась.

На этом, ребята, до свидания. Всего самого-самого хорошего вам. И ещё вот что, староста, Дима, сообщи тем, кто отсутствовал, вот то, что я рассказывал, ну, вкратце хотя бы. Ну хотя бы ссылку дай на задание и скажи, что без этого, так сказать, они пятёрки не получат.

Хорошо, я им скажу об этом.

Ну, если там, допустим, на красный диплом идёт, и все у него пятёрки, ну я не поставлю четвёрку, конечно. Вот. Но вообще-то вот пятёрку ставят за работу реальную, за то, что человек продемонстрировал свои знания, умения, навыки и компетенции. Это делается вот таким способом в данном случае. Понятно, да? Всего самого хорошего, ребята. До свидания.

До свидания. До свидания.

До свидания.