

# Аналитическая платформа PolyAnalyst 6.5



Рабочая программа курса

«Работа с платформой PolyAnalyst»

# ПРОГРАММА

**Наименование программы:** Работа с платформой PolyAnalyst.

**Количество часов:** 19 занятий по 90 минут = 40 академических часов.

**Формат обучения:** онлайн форма обучения.

**Материально-техническое оснащение:** для выполнения практических заданий необходимы компьютеры (ноутбуки) с операционной системой Windows 8.1/10, выход в интернет, микрофон (вебкамера с микрофоном), Zoom.

Осваиваемые компетенции	Практические задания	Узлы	Количество часов (академических)
<b>Занятие 0 (Установочное)</b>			1
<ul style="list-style-type: none"><li>Регистрация и подключение к серверу PolyAnalyst</li><li>Использование справочной информации</li></ul>	<ul style="list-style-type: none"><li>Настройка подключения к учебному серверу</li><li>Поиск по справке</li></ul>		
<b>Занятие 1 «Знакомство с платформой PolyAnalyst»</b>			2
<ul style="list-style-type: none"><li>Запуск программы</li><li>Операции с проектом (создание, импорт, экспорт)</li><li>Знакомство с рабочими панелями аналитика</li><li>Знакомство с базовыми узлами импорта данных</li><li>Знакомство с базовыми действиями в платформе PolyAnalyst (добавление, копирование и удаление узлов, создание «скрипта»)</li><li>Типы данных в PolyAnalyst</li><li>Просмотр табличных данных, базовая статистика</li></ul>	<ul style="list-style-type: none"><li>Запуск PolyAnalyst</li><li>Создание нового проекта</li><li>Создание простого скрипта, состоящего из нескольких «узлов»</li><li>Экспорт/импорт проекта</li><li>Одновременная работа с несколькими проектами</li><li>Копирование узлов в другой проект</li></ul>	<ul style="list-style-type: none"><li>Файлы CSV</li></ul>	
<b>Занятие 2 «Импорт данных»</b>			2
<ul style="list-style-type: none"><li>Анализ ошибок импорта данных</li><li>Импорт данных из базы данных через ODBC</li><li>Импорт данных из источников в сети Интернет</li><li>Импорт электронной почты</li><li>Использование OCR для распознавания документов</li><li>Импорт данных из SharePoint</li></ul>	<ul style="list-style-type: none"><li>Устранение ошибок при импорте данных</li><li>Импорт данных из базы данных через ODBC</li><li>Импорт данных из источников в сети Интернет</li><li>Настройка OCR для импорта данных из изображений</li><li>Импорт данных из JSON или XML</li></ul>	<ul style="list-style-type: none"><li>Microsoft Excel</li><li>Файлы / Изображения / FTP</li><li>ODBC / Microsoft Access</li><li>E-Mail / E-Mail Архив</li><li>Интернет / RSS</li><li>SharePoint</li><li>XML / JSON</li></ul>	
<b>Занятие 3 «Базовые операции с данными»</b>			2
<ul style="list-style-type: none"><li>Базовые операции с таблицей (переименование, удаление столбцов)</li><li>Фильтрация данных</li><li>Сортировка</li><li>Удаление дубликатов</li><li>Создание выборки</li></ul>	<ul style="list-style-type: none"><li>Переименование/удаление/изменение типа данных колонки</li><li>Фильтрация данных по параметрам</li><li>Создание выборки из генеральной совокупности</li><li>Просмотр и сравнение базовой статистики для выборки и генеральной совокупности</li></ul>	<ul style="list-style-type: none"><li>Фильтрация колонок</li><li>Модификация колонок</li><li>Фильтрация строк</li><li>Сортировка строк</li><li>Уникальные записи</li><li>Выборка</li></ul>	

<ul style="list-style-type: none"> <li>• Создание новой колонки</li> <li>• Функции расчета значений новой колонки</li> <li>• Язык SRL, базовые операции, группы функций</li> </ul>	<ul style="list-style-type: none"> <li>• Добавление новой колонки к существующим данным</li> <li>• Использование логических, арифметических и условных функций для заполнения колонки</li> <li>• Использование функций языка SRL для заполнения колонки</li> </ul>	<ul style="list-style-type: none"> <li>• Производные колонки</li> </ul>
--	--	---

## Занятие 4 «Дополнительные операции с данными. Экспорт данных» 2

<ul style="list-style-type: none"> <li>• Агрегирование данных</li> <li>• Объединение таблиц</li> <li>• Конкатенация таблиц</li> <li>• Дополнение таблиц</li> </ul>	<ul style="list-style-type: none"> <li>• Вычисление агрегаций для набора данных</li> <li>• Объединение данных из двух таблиц</li> <li>• Конкатенация данных из двух таблиц</li> <li>• Вычисление разницы для двух таблиц</li> <li>• Сравнение двух таблиц</li> </ul>	<ul style="list-style-type: none"> <li>• Агрегирование</li> <li>• Объединение</li> <li>• Конкатенация</li> <li>• Сравнение</li> <li>• Дополнение</li> </ul>
<ul style="list-style-type: none"> <li>• Экспорт данных в файл / папку</li> <li>• Экспорт данных в ODBC</li> <li>• Экспорт данных в Microsoft Word</li> <li>• Отправка писем</li> </ul>	<ul style="list-style-type: none"> <li>• Настройка экспорта в БД через ODBC</li> <li>• Экспорт в Microsoft Word с использованием шаблона слияния</li> </ul>	<ul style="list-style-type: none"> <li>• Экспорт в файл</li> <li>• Экспорт в папку</li> <li>• Экспорт в ODBC</li> <li>• Экспорт в Microsoft Word</li> <li>• Отправка писем</li> </ul>

## Занятие 5 «Введение в анализ данных. Классификация» 2

<ul style="list-style-type: none"> <li>• Обучение с учителем и без учителя, проблема переобучения</li> <li>• Данные для обучения и валидации модели</li> </ul>	<ul style="list-style-type: none"> <li>• Разделение данных на данные для обучения и валидации модели</li> </ul>	<ul style="list-style-type: none"> <li>• Разбиение</li> </ul>
<ul style="list-style-type: none"> <li>• Задача классификации</li> <li>• Подходы к решению задачи классификации (логистическая регрессия, линейная классификация, дерево решений, случайный лес, SVM)</li> <li>• Оценка результатов классификации (матрица ошибок)</li> </ul>	<ul style="list-style-type: none"> <li>• Построение моделей классификации при помощи разных алгоритмов (минимум двух)</li> </ul>	<ul style="list-style-type: none"> <li>• Логистическая регрессия</li> <li>• Линейная классификация</li> <li>• Дерево решений</li> <li>• Случайный лес</li> <li>• SVM-классификатор</li> </ul>
<ul style="list-style-type: none"> <li>• Применение модели к новым данным</li> <li>• Валидация моделей</li> <li>• Сравнение и выбор наилучшей модели</li> </ul>	<ul style="list-style-type: none"> <li>• Сравнение двух моделей классификации</li> <li>• Выбор оптимальной модели</li> <li>• Применение модели к новым данным</li> </ul>	<ul style="list-style-type: none"> <li>• Объединение моделей</li> <li>• Тестирование моделей</li> <li>• Применение моделей</li> </ul>

## Занятие 6 «Регрессионный анализ. Кластеризация. Нейронные сети» 2

<ul style="list-style-type: none"> <li>• Задача регрессионного анализа</li> <li>• Оценка результатов регрессионного анализа</li> </ul>	<ul style="list-style-type: none"> <li>• Выбор переменных для регрессионной модели на основе матрицы диаграмм рассеивания</li> <li>• Построение модели линейной регрессии</li> </ul>	<ul style="list-style-type: none"> <li>• Матрица диаграмм рассеивания</li> <li>• Линейная регрессия</li> <li>• Многомерные адаптивные регрессионные сплайны</li> </ul>
<ul style="list-style-type: none"> <li>• Задача кластеризации</li> <li>• Кластеризация методом К-средних</li> <li>• Метод максимизации ожидания</li> <li>• Кластеризация на основе расстояния</li> <li>• Сеть Кохонена</li> </ul>	<ul style="list-style-type: none"> <li>• Разбиение данных на кластеры с использованием разных алгоритмов (не менее двух)</li> </ul>	<ul style="list-style-type: none"> <li>• Кластеризация по методу k-средних</li> <li>• Максимизация ожидания</li> <li>• Кластеризация на основе расстояния</li> <li>• Сеть Кохонена</li> </ul>
<ul style="list-style-type: none"> <li>• Классическая нейронная сеть (базовые понятия, выбор оптимальной структуры сети)</li> <li>• Сверточная нейронная сеть, отличия от классической нейронной сети</li> </ul>	<ul style="list-style-type: none"> <li>• Построение нейронной сети для задачи классификации</li> </ul>	<ul style="list-style-type: none"> <li>• Нейронная сеть</li> <li>• Сверточная нейронная сеть</li> </ul>

## Занятие 7 «Основы обработки текстовых данных. Словари»

2

- Основы обработки текстовых данных
- Определение языка, индексирование, проверка орфографии, извлечение ключевых слов
- Удаление однотипных фрагментов текста
- Удаление похожих документов
- Определение языка и построение индекса для корпуса текстов
- Проверка и исправление орфографии
- Построение списка ключевых слов
- Удаление дублирующихся текстов и повторяющихся фрагментов
- Определение языка
- Производная таблица
- Индекс
- Проверка орфографии
- Удаление фрагментов
- Уникальные тексты
- Извлечение ключевых слов
- Типы и виды словарей, их назначение
- Менеджер словарей, операции со словарями
- Операции со словарями (создание, редактирование, сравнение, импорт, экспорт)
- Экспорт в словарь

## Занятие 8 «Извлечение стандартных сущностей.

2

### Анонимизация и восстановление сущностей»

- Извлечение стандартных сущностей (имена людей, названия компаний, организаций, географические объекты и тд)
- Извлечение фактов
- Анализ тональности
- Извлечение стандартных именованных сущностей и фактов из текста
- Отображение связей между сущностями
- Извлечение сущностей
- Извлечение фактов
- Анализ тональности
- Производная таблица
- Анонимизация и восстановление сущностей
- Анонимизация и восстановление сущностей в документах
- Анонимизация сущностей
- Восстановление сущностей

## Занятие 9 «Классификация и кластеризация текстов»

2

- Определение тематики текста
- Классификация и кластеризация текстов
- Настройки предикторов (лексемы, ключевые слова, сущности)
- Обучение классификатора
- Применение обученной модели к новому набору данных
- Тематика текстов
- Классификация текстов
- Кластеризация текстов

## Занятие 10 «Язык поисковых запросов PDL»

2

- Создание поисковых запросов PDL, в т.ч. с использованием:
  - операторов PDL: or, and, not;
  - функций работы с морфологией (lemma, form);
  - функций поиска последовательностей (sentence, optional, phrase, near, repeat) именованных параметров функций;
  - функций работы с пунктуацией;
  - функций обращения к словарю;
  - функций поиска стандартных сущностей.
- Написание запросов на языке PDL для поиска информации в текстах
- Поисковый запрос

## Занятие 11 «Таксономия. Классификация текстов на основе правил»

2

- Классификация текстов на основе правил
- Создание таксономии для классификации документов на основе правил
- Таксономия
- Параметры

## Занятие 12 «Язык извлечения информации XPDL»

2

- Знакомство с языком XPDL (XPDL и PDL)
- Структура правил языка XPDL
- Работа с редактором правил
- Понятие пользовательской сущности
- Добавление пользовательской сущности в узел Извлечение сущностей
- Создание правила на языке XPDL
- Отображение результатов
- Извлечение сущностей

<b>Занятие 13 «Дополнительные возможности языка XPDL»</b>		<b>2</b>
<ul style="list-style-type: none"> <li>• Вложенные правила</li> <li>• Форматирующие функции</li> <li>• Постобработка</li> <li>• Использование ранее извлеченных сущностей</li> </ul>	<ul style="list-style-type: none"> <li>• Создание набора правил на языке XPDL с использованием вложенных правил</li> <li>• Нормализация и форматирование извлеченной информации</li> </ul>	<ul style="list-style-type: none"> <li>• Создание набора правил на языке XPDL с использованием вложенных правил</li> <li>• Нормализация и форматирование извлеченной информации</li> </ul>
<b>Занятие 14 «Самостоятельная работа "Разработка решения для анализа текстов на естественном языке"»</b>		<b>3</b>
	<p>Создание решения включающего следующие этапы:</p> <ul style="list-style-type: none"> <li>• импорт данных в систему</li> <li>• предобработка текстовых данных</li> <li>• создание правила для извлечения пользовательских сущностей из текстов</li> <li>• постобработка результатов</li> </ul>	<ul style="list-style-type: none"> <li>• Извлечение сущностей</li> <li>• Производная таблица</li> </ul>
<b>Занятие 15 «Основы администрирования сервера PolyAnalyst»</b>		<b>2</b>
<ul style="list-style-type: none"> <li>• Модель разграничения прав в PolyAnalyst на основе пользователей и групп</li> <li>• Типы пользователей (Regular, LDAP, OAuth, Guest)</li> <li>• Создание, удаление и управление пользователями</li> <li>• Форма регистрации пользователей</li> <li>• Требования к паролю и настройки безопасности</li> <li>• Папки пользователя. PolyAnalyst Drive</li> </ul>	<ul style="list-style-type: none"> <li>• Создание пользователей</li> <li>• Создание формы для регистрации пользователей</li> <li>• Назначения лицензий и прав пользователям</li> <li>• Сброс/восстановление пароля пользователя</li> <li>• Настройка LDAP-сервера для авторизации</li> </ul>	
<ul style="list-style-type: none"> <li>• Права доступа на проект/отчет</li> <li>• Понятие владлец проекта/отчета. Операции разрешенные только для владельца</li> <li>• Права доступа на отдельную колонку (персональные данные)</li> <li>• Права доступа на словари</li> </ul>	<ul style="list-style-type: none"> <li>• Настройка прав пользователей на проект</li> <li>• Настройка прав на отдельную колонку таблицы</li> </ul>	
<ul style="list-style-type: none"> <li>• Настройка автоматического обновления</li> <li>• Downloadable content</li> </ul>	<ul style="list-style-type: none"> <li>• Настройка автоматического обновления</li> <li>• Настройка пользовательского источника DLC</li> </ul>	
<b>Занятие 16 «Планировщик заданий PolyAnalyst. REST API»</b>		<b>2</b>
<ul style="list-style-type: none"> <li>• Виды событий планировщика</li> <li>• Типы и параметры задач планировщика</li> <li>• Импорт и экспорт задач планировщика</li> </ul>	<ul style="list-style-type: none"> <li>• Настройка задачи в планировщике для выполнения по расписанию</li> </ul>	
<ul style="list-style-type: none"> <li>• Доверенные приложения</li> </ul>	<ul style="list-style-type: none"> <li>• Настройка доверенного приложения</li> </ul>	
<ul style="list-style-type: none"> <li>• Автоматизация задач с помощью REST API</li> <li>• Группы операций, доступных через REST API</li> </ul>	<ul style="list-style-type: none"> <li>• Использование curl для выполнения REST-запросов</li> </ul>	
<b>Занятие 17 «Использование Python для расширения возможностей системы»</b>		<b>2</b>
<ul style="list-style-type: none"> <li>• Доступные SDK для REST API</li> </ul>	<ul style="list-style-type: none"> <li>• Создание скрипта с помощью Python SDK для выполнения автоматизированных задач</li> </ul>	
<ul style="list-style-type: none"> <li>• Интерпретаторы для узлов R и Python</li> </ul>	<ul style="list-style-type: none"> <li>• Настройка интерпретатора для узлов Python/R</li> </ul>	<ul style="list-style-type: none"> <li>• Python</li> <li>• R</li> </ul>
<ul style="list-style-type: none"> <li>• Веб-скраперы</li> </ul>	<ul style="list-style-type: none"> <li>• Создание и подключение пользовательского веб-скрапера</li> </ul>	<ul style="list-style-type: none"> <li>• Интернет</li> </ul>

<b>Занятие 18 «Визуализация результатов»</b>		<b>2</b>
<ul style="list-style-type: none"> <li>• Знакомство с базовыми способами визуализации информации в PolyAnalyst</li> </ul>	<ul style="list-style-type: none"> <li>• Построение базовых объектов визуализации</li> </ul>	<ul style="list-style-type: none"> <li>• Агрегированный график</li> <li>• Круговая диаграмма</li> <li>• Диаграмма рассеивания</li> <li>• Линейный график</li> <li>• Пузырьковая диаграмма</li> <li>• Облако тегов</li> </ul>
<ul style="list-style-type: none"> <li>• Знакомство со средствами визуализации данных в виде графов в системе PolyAnalyst</li> </ul>	<ul style="list-style-type: none"> <li>• Создание узлов для визуализации данных в виде графа</li> </ul>	<ul style="list-style-type: none"> <li>• Диаграмма связей</li> <li>• Граф</li> </ul>
<ul style="list-style-type: none"> <li>• Знакомство с узлами многомерного анализа</li> </ul>	<ul style="list-style-type: none"> <li>• Создание и работа с узлами многомерного анализа</li> </ul>	<ul style="list-style-type: none"> <li>• Многомерная матрица</li> <li>• OLAP таблица</li> </ul>
<ul style="list-style-type: none"> <li>• Знакомство с узлами ГИС</li> </ul>	<ul style="list-style-type: none"> <li>• Наложение данных на карту</li> </ul>	<ul style="list-style-type: none"> <li>• Карта</li> <li>• Объект GIS</li> </ul>

<b>Занятие 19 «Интерактивные веб-отчеты»</b>		<b>2</b>
<ul style="list-style-type: none"> <li>• Создание веб-отчета</li> <li>• Базовые компоненты отчета</li> <li>• Добавление компонентов в веб-отчет</li> <li>• Настройка связей между компонентами отчета</li> <li>• Работа с публикациями</li> </ul>	<ul style="list-style-type: none"> <li>• Создание веб-отчета</li> <li>• Публикация веб-отчета</li> </ul>	
<ul style="list-style-type: none"> <li>• Отчеты, доступные публично</li> <li>• Встраивание отчетов во внешний сайт</li> </ul>	<ul style="list-style-type: none"> <li>• Встраивание отчета во внешний сайт</li> </ul>	

**Итого** **40**