

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ АЛГОРИТМИЗАЦИЯ И ПРОГРАММИРОВАНИЕ РАСПРЕДЕЛЕННЫХ ДАННЫХ

Автор: Стефановский Дмитрий Владимирович

Код и наименование направления подготовки, профиля: 38.04.01 Экономика («Системы больших данных в экономике»)

Квалификация (степень) выпускника: Магистр

Форма обучения: очная

Цель освоения дисциплины:

Сформировать компетенции в сфере аналитической работы в области экономики и финансов, обработки и анализов данных.

План курса:

Тема 1. Предварительные сведения о языке Питон.

Почему именно Питон. Питон как клей. Решение проблемы «двух языков». Недостатки Питон. Необходимые библиотеки для Питон. Установка и настройка. Сообщество и конференции.

Тема 2. Интерактивные вычисления и среда разработки.

Основы Питон. История команд. Взаимодействие с операционной системой. Средства разработки программ. HTML-блокнот в IPython. Советы по продуктивной разработке кода с использованием IPython.

Тема 3. Переформатирование данных: очистка, преобразование, слияние, изменение формы

Комбинирование и слияние наборов данных. Изменение формы и поворот. Преобразование данных. Манипуляции со строками. Пример: база данных реальных рожд в сетях магазинов.

Тема 4. Построение графиков и визуализация. Агрегирование данных и групповые операции. Временные ряды

Краткое введение в API библиотеки matplotlib. Функции построения графиков в pandas. Нанесение данных на карту. Инструментальная экосистема визуализации для Питон. Механизм GroupBy. Агрегирование данных. Групповые операции и преобразования. Сводные таблицы и кросс-табуляция. Типы данных и инструменты, относящиеся к дате и времени. Основы работы с временными рядами. Диапазоны дат, частоты и сдвиг. Часовые пояса. Передискретизация и преобразование частоты. Графики временных рядов. Скользящие оконные функции.

Тема 5. Предмет распределенных вычислений

Понятия распределенных вычислений и распределенной системы. Цели построения распределенных систем. Требования к распределенным системам: прозрачность, открытость, масштабируемость, сложность разработки распределенных систем. Понятие и назначение программного обеспечения промежуточного уровня. Взаимодействие в распределенных системах: физическое время, синхронные и асинхронные распределенные системы, упорядочивание событий, примитивы взаимодействия. Синхронный и асинхронный обмен сообщениями.

Тема 6 Модель распределенного вычисления

Модель распределенной системы. Причинно-следственный порядок событий. Эквивалентные выполнения. Конус прошлого и конус будущего для события. Свойства каналов.

Тема 7. Логические часы

Общие принципы построения логических часов. Скалярное время Лэмпорта: основные свойства, примеры использования. Векторное время: основные свойства, примеры использования. Методы эффективной реализации векторных часов: дифференциальная пересылка векторного времени; часы, фиксирующие прямую зависимость; адаптивный метод Жарда-Жордана. Матричное время, основные свойства.

Тема 8. Взаимное исключение в распределенных системах.

Общие концепции. Централизованный алгоритм. Алгоритмы на основе получения разрешений: алгоритм Лэмпорта, алгоритм Рикарта-Агравала, алгоритм обедающих философов. Алгоритмы на основе передачи маркера: широковещательный алгоритм Сузуки-касами, алгоритм Реймонда на основе покрывающего дерева.

Аудиторные часы: 144

Формы текущего контроля и промежуточной аттестации: опросы, домашнее задание, контрольная работа, экзамен.

Основная литература:

1. Уэс Маккинли. Python и анализ данных/ Пер. с англ. Слинкин А. А. - М.: ДМК Пресс, 2015. - 482 с.: ил.
2. Косяков М.С. Введение в распределенные вычисления. – СПб: НИУ ИТМО, 2014. – 155 с.