

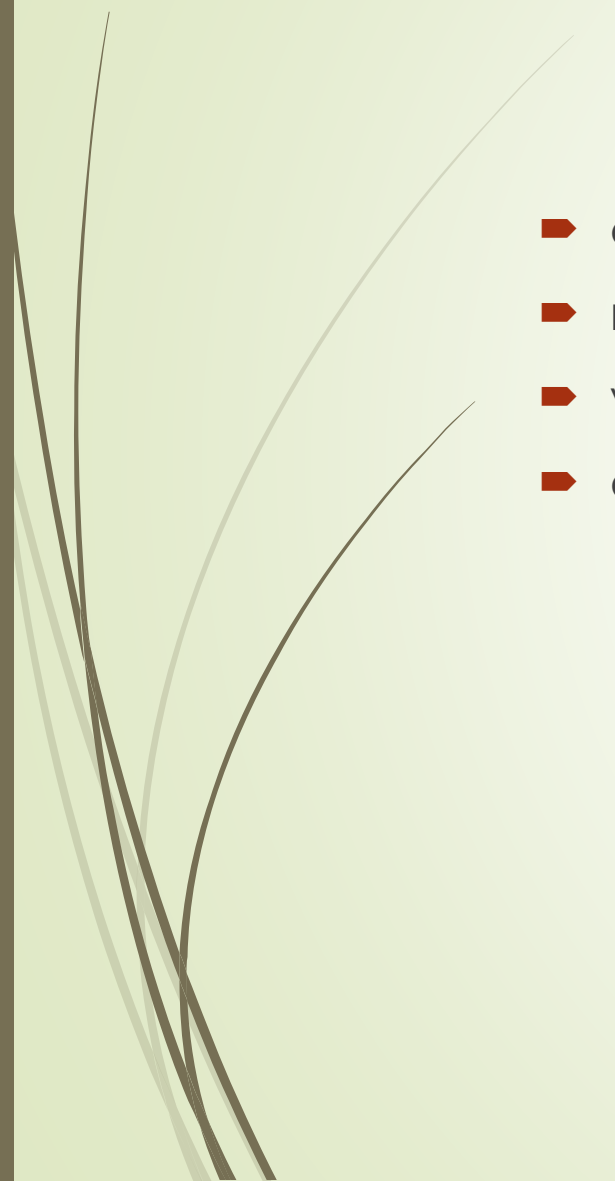


Статистические ПОИС. Архитектура СПОИС.

Информационное обеспечение предметно-ориентированных ИС.



Типовые задачи СПОИС

- сбор и агрегация данных
 - процедуры очистки, проверки и фильтрации данных
 - упорядочение и хранение статистической информации
 - анализ и обработка хранимой статистической информации
- 



Типовой функционал СПОИС на примере SAS

- Менеджмент статистической информации
- Data mining
- Data science
- Прогнозирование и эконометрика
- Высокопроизводительная аналитика (Big Data)
- Статистический анализ
- Симуляции
- Интеллектуальный анализ текста

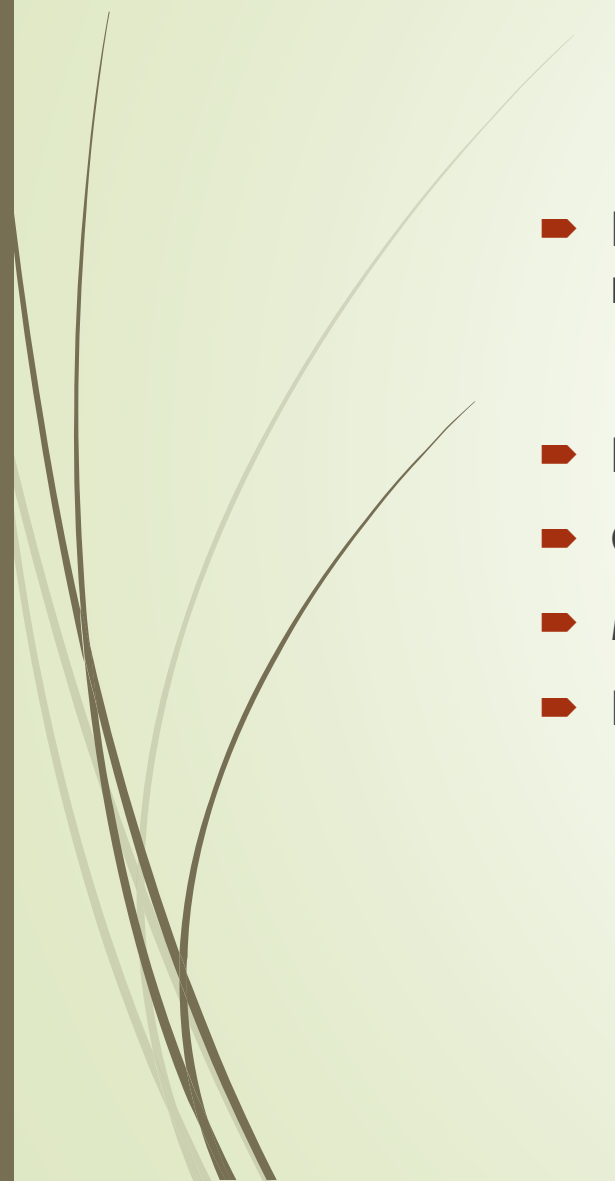


Data mining

- Совокупность методов обнаружения в данных ранее неизвестных, практически полезных знаний, необходимых для принятия решений.
- Ассоциативные правила или паттерны
- Группировка объектов, кластерный анализ
- Построение модели регрессии
- Классификация объектов
- Регрессионный анализ (временные ряды)



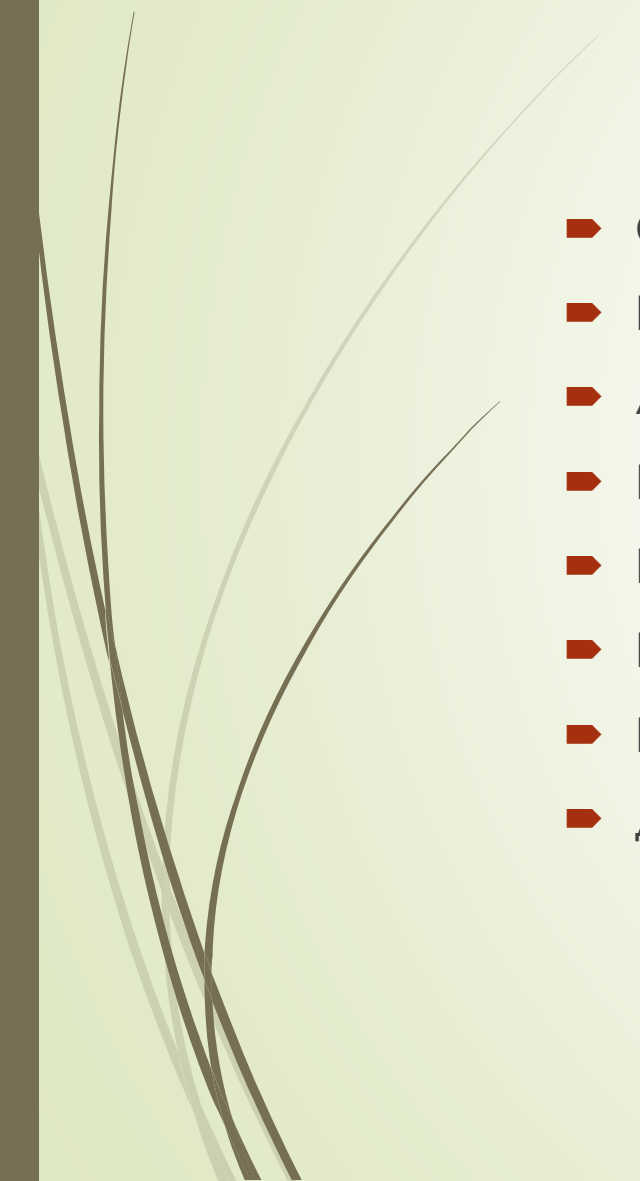
Data Science



- Раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме.
- Проектирование алгоритмов mapReduce
- Семплирование и регрессии
- Машинное обучение
- Вероятностные модели



Статистический анализ

- Статистические наблюдения
 - Группировки
 - Абсолютные и относительные статистические величины
 - Вариационные ряды
 - Выборка
 - Корреляционный анализ
 - Регрессионный анализ
 - Динамические ряды
- 



Интеллектуальный анализ текста

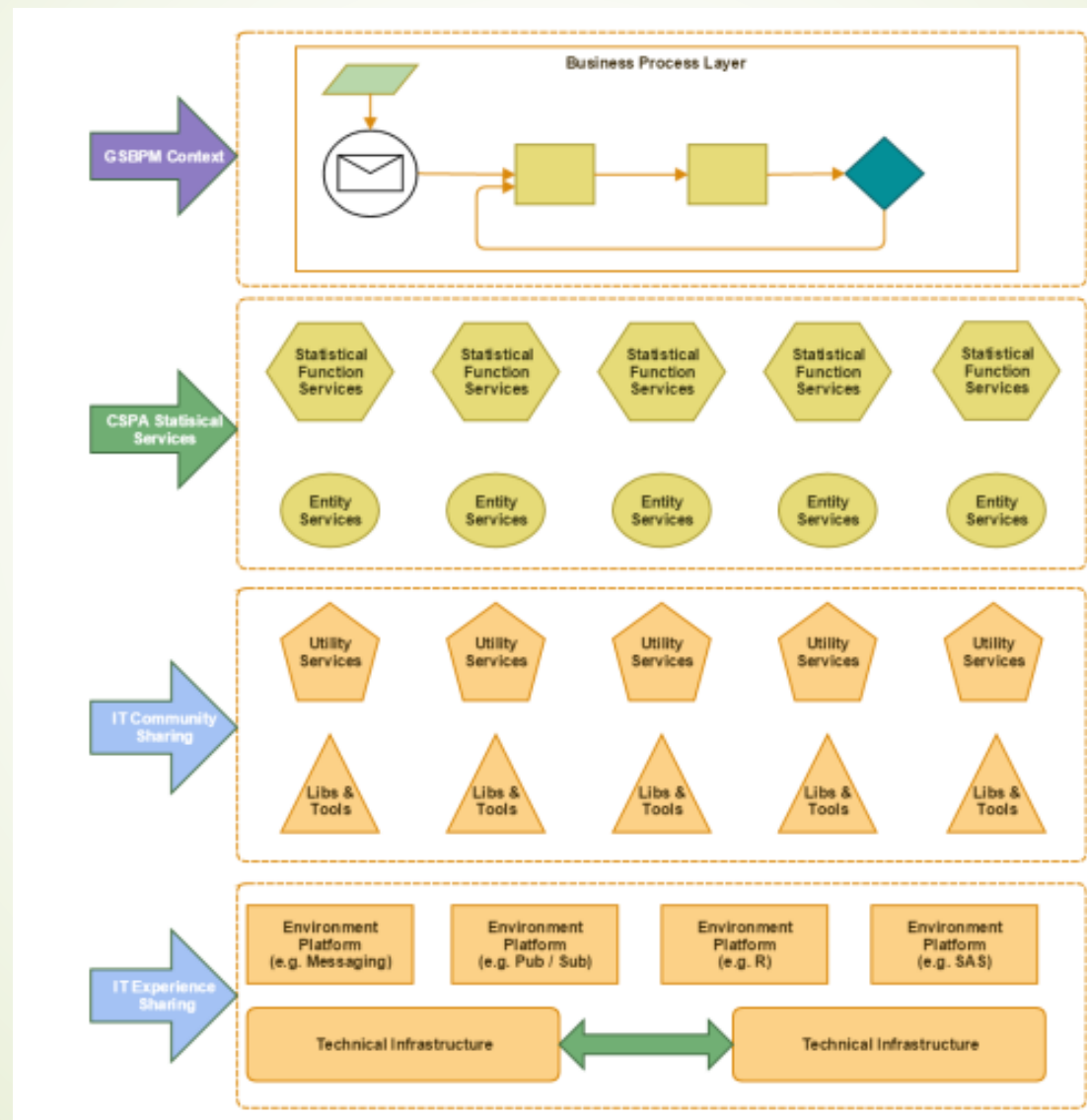
- Направление в искусственном интеллекте, целью которого является получение информации из коллекций текстовых документов, основываясь на применении эффективных в практическом плане методов машинного обучения и обработки естественного языка.
- Категоризация текстов
- Извлечение информации
- Обработка изменений в коллекциях
- Средства представления информации для пользователя



Общая архитектура производства статистической информации (CSPA)

- Модель, предназначенная для статистической индустрии. Разработана международным сообществом статистиков. Основные области применения:
- Производство статистической информации.
- Обеспечение GSIM
- Архитектура приложений
- Технологическая архитектура

Архитектура приложений CSPА

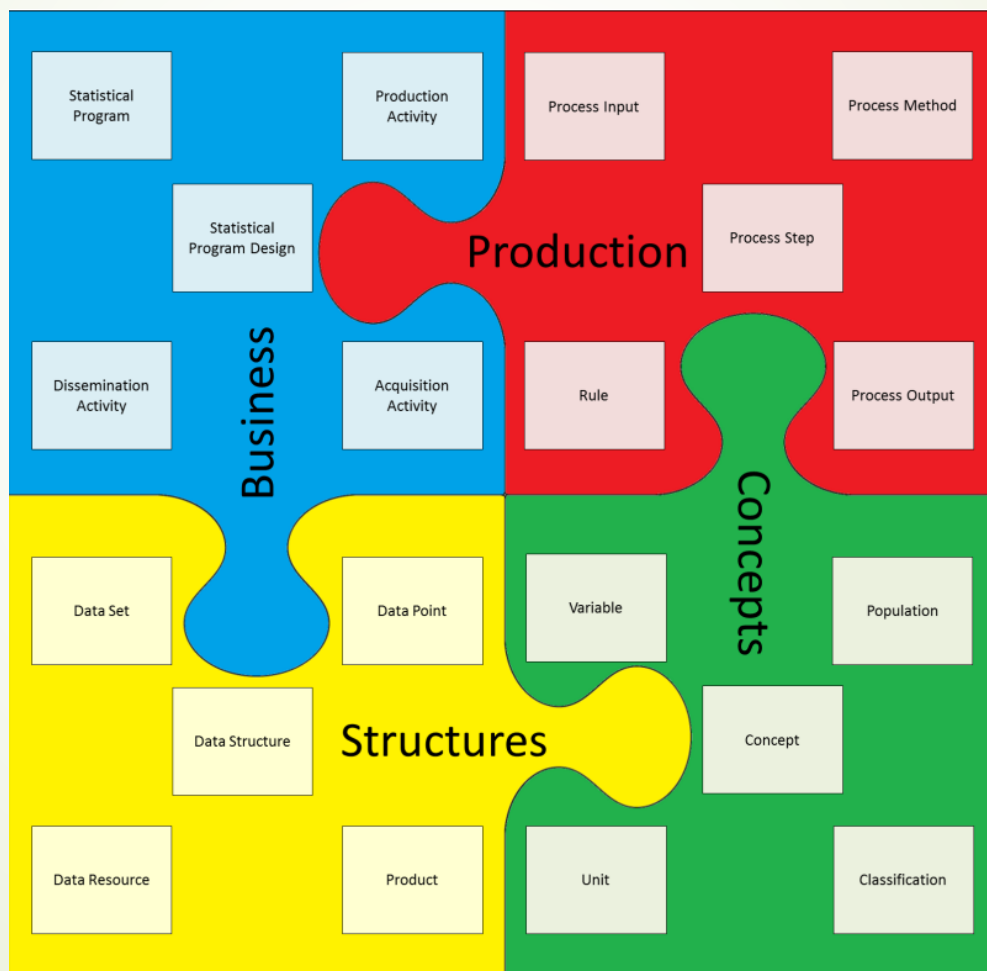




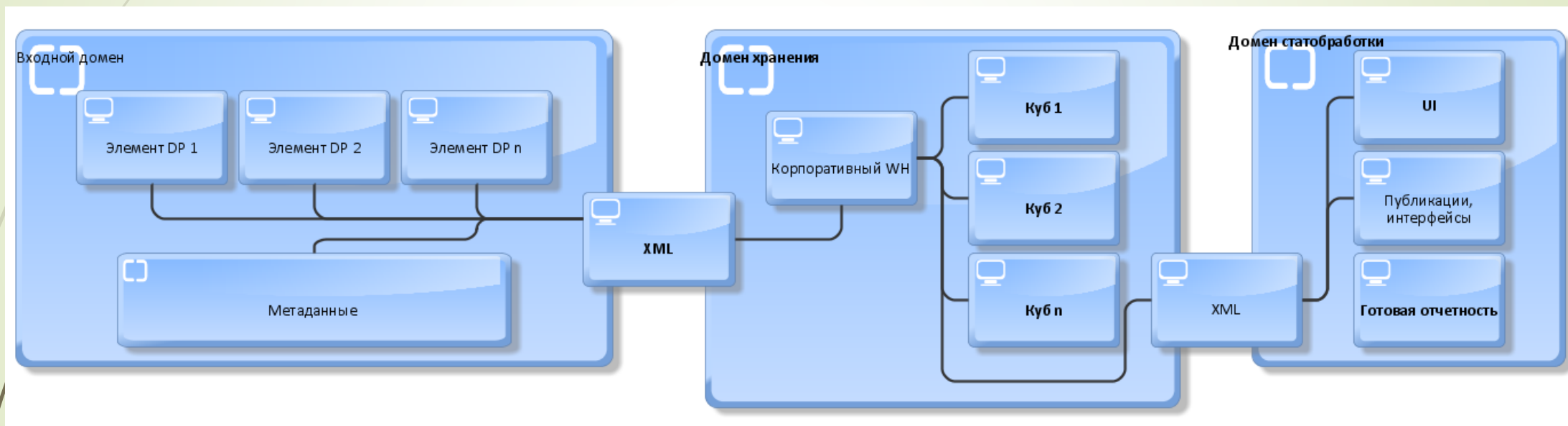
Общая статистическая информационная модель (GSIM)

- Фреймворк сущностей, атрибутов и связей, который описывает информацию, используемую при производстве официальной статистики (информационные объекты). В общем виде состоит из 4-ех взаимосвязанных элементов:
- Структура
- Идея
- Производство
- Бизнес

Общая статистическая информационная модель (GSIM)



Образец. Архитектура OECD (SIS)



Архитектура SAS (с точки зрения интероперабельности)

