

Проектирование и разработка информационной системы с использованием динамических баз знаний

РЯБОВ Д.В.

База знаний в информационной системе

Разработка информационной системы с использованием динамической базы знаний, построенной на потоке данных социальных сетей

На основе полученных данные возможно сформировать динамическую базу знаний в виде сетевой модели.

БЗ станет центральным компонентом разрабатываемой информационной системы

Динамическая БЗ - модификация БЗ в процессе работы интеллектуальной системы

Данные, получаемые из социальных сетей, слабоструктурированные что затрудняет их обработку

Основные свойства данных — разнообразие, высокая скорость поступления и большой объем

Хранение данных (NoSQL базы данных)

Масштабируемость

- Горизонтальное масштабирование
- Распределенные системы

Слабо типизированная структура

- Хранение неструктурированной информации
- Хранение гетерогенных данных

Эффективной работы с разреженными данными

MongoDB

Документированная БД (единицей хранения данных является документ – некоторый объект, обладающий произвольным набором атрибутов)

Гибкая структура хранения данных

JSON (JavaScript Object Notation)

Neo4J

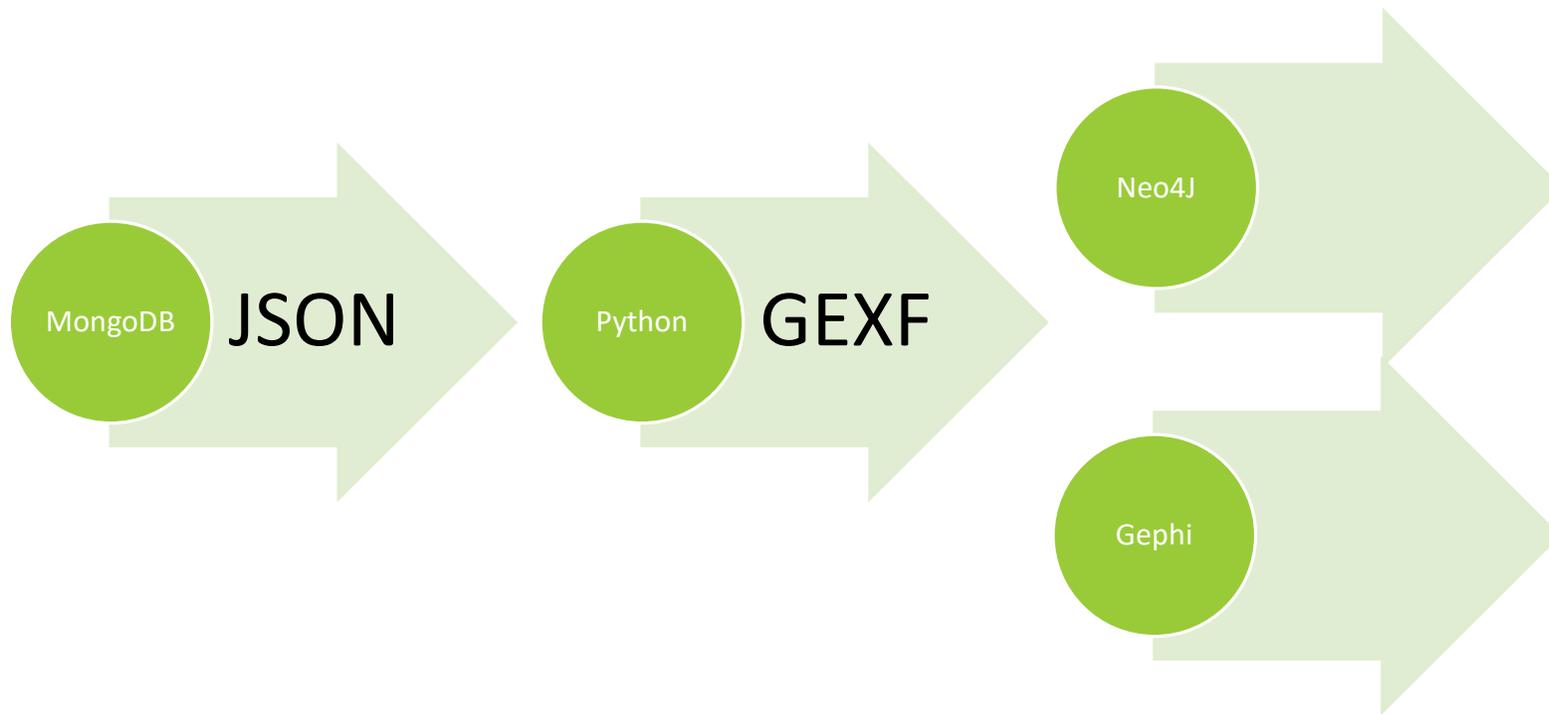
Подходят для реализации проектов, которые предполагают изначально естественную графовую структуру данных

- Операции на графами

Оперирует понятиями: вершина, связь, свойства, метки

Построение семантической сети

Эксперимент



Тематическое моделирование

Тематический поиск

Ранжирование документов по степени тематического сходства

Построение иерархический тематический каталог коллекции документов

Определение изменения темы во времени

Определение тематики автора

Определение тематики различных сущностей, связанных с документами

Вероятностные методы

pLSA - Probabilistic Latent Semantic Analysis. Вероятностный латентно-семантический анализ

LDA - Latent Dirichlet Allocation. Латентное размещение Дирихле

ARTM - Additive Regularization of Topic Models. Аддитивная Регуляризация Тематических Моделей

Нейронные сети

Top2Vec - алгоритм тематического моделирования и семантического поиска, который автоматически определяет темы, присутствующие в тексте

LDA2Vec – гибридный алгоритм LDA + Word2vec

Спасибо за внимание!

