

Вопросы

- Основные этапы извлечения и интеграции информации
- Примеры приложений и проектов для интеграции информации
- Специфика извлечения и интеграции информации из больших данных
- Виды источников данных
- Задачи курирования данных
- Проблемы использования открытых данных и подходы к их решению
- Что такое извлечение информации из текстовых данных
- Виды представления извлеченных данных
- Примеры приложений и проектов, использующих извлечение информации из текстов
- Способы извлечения информации из текстов
- Что такое неоднозначность понятий и подходы к разрешению неоднозначностей
- Процесс интеграции информации. Сопоставление схем. Трансформация данных
- Подходы к разрешению конфликтов схем
- Разрешение сущностей (Entity Resolution, ER). Выявление дубликатов, удаление дубликатов, установление связей между сущностями из разных исходных коллекций
- Алгоритмы ER. Data Preparation and Match Features
- Алгоритмы ER. Pairwise ER
- Алгоритмы ER. Constraints in ER
- Алгоритмы ER. Record linkage, Deduplication, Collective
- Масштабирование ER для больших данных. Blocking. Distributed ER
- Relational ER. Sorted Neighborhood Method
- Hierarchical ER. Containment Metric. Top-Down & Bottom-Up Algorithms
- Graph ER. Graph Entity Resolution Framework
- Что такое Слияние данных (Data Fusion, DF)
- Слияние данных. Completeness, Conciseness, and Correctness
- Слияние данных. Стратегии разрешения конфликтов. Функции разрешения конфликтов
- Слияние данных. Union-based Approaches. Union, Outer Union, Minimum Union, Complement Union
- Слияние данных. Join-based Approaches. Full Disjunction, Match Join, Grouping And Aggregation
-