TECHNISCHE UNIVERSITÄT DRESDEN

SEMINAR SELECTED TOPICS IN DATABASE THEORY

Lecture 2: Literature

David Carral, Markus Krötzsch

Knowledge-Based Systems

TU Dresden, 10th October 18

Evaluation

Paper Summary

- Self-selected research paper (see the “Literature” tab at `https://iccl.inf.tu-dresden.de/web/Research_Advances_in_Database_Theory_(WS2018)`)
- ~10 pages

Presentation

- 20 minutes + discussion
- Participate in the presentations of other students

Remark: Check the examination requirements!

Document Spanners:
A Formal Approach to Information Extraction

"An intrinsic part of information extraction is the creation and manipulation of relations extracted from text. In this paper, we develop a foundational framework where the central construct is what we call a spanner. A spanner maps an input string into relations over the spans (intervals specified by bounding indices) of the string. The focus of this paper is on the representation of spanners."

- Information Retrieval.
- Pattern Recognition.
- Mathematical Logic and Formal Languages.

Optimal Aggregation Algorithms for Middleware

Development of an algorithm to rank top k objects in a database with high overall grade using a function over a set of attributes.

- Algorithmic Complexity.
- Probability.

Why and Where: A Characterization of Data Provenance

Data provenance: Establishing where a piece of data came from and the process by which it arrived in the database.

Contribution: Describe an approach to compute provenance when the data of interest has been created by a database query.
Two different types of provenance: “why” provenance (refers to the source data that had some influence on the existence of the data) and “where” provenance (refers to the location(s) in the source databases from which the data was extracted).

- Query answering databases.
- Provenance.

Counting Beyond a Yottabyte, or how SPARQL 1.1 Property Paths will Prevent Adoption of the Standard

SPARQL is an RDF query language—that is, a semantic query language for databases—able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

Contribution: This paper discusses the evaluation methods for regular path queries in SPARQL which often exhibit poor performance. Moreover, the authors succesfully argue that this poor performance is due to a “bug” in the design of the language semantics and propose a suitable “fix.”

- Semantic Web.
- Regular path queries.

A Formal Semantics of SQL Queries, Its Validation, and Applications

Contribution: Definition of semantics for SELECT-FROM-WHERE queries without any departures from the real languages. This fragment already requires decisions related to the data model and handling variable names that are normally disregarded by simplified semantics. This new semantics allow use to deal with bag semantics and nulls.

- Conjunctive query answering.
- Formal Semantics.

Parallel-Correctness and Transferability for Conjunctive Queries

Contribution: Introduction of a condition, called parallel-correctness, for the evaluation of queries w.r.t. a distribution policy. Furthermore, the authors study the complexity of parallel-correctness for conjunctive queries as well as transferability of parallel-correctness between queries.

- Conjunctive query answering.
- Complexity Theory.
- Parallel Computation.

Data Exchange: Semantics and Query Answering

"In this paper, we address foundational and algorithmic issues related to the semantics fo data exchange and to the query answering problem in the context of data exchange. We study the problem of computing the certain answers of target queries by simply evaluating them on a canonical universal solution, and we explore the boundary of what queries can and cannot be answered this way, in a data exchange setting."

- Reasoning.
- First-order (in)expressibility.
- Complexity.
- Query Answering.

Querying Graphs with Data

"In this paper we present a comprehensive study of languages that enable such combination of data and topology querying. These languages come in two flavors. The first follows the standard approach of path queries, which specify how labels of edges change along a path, but now we extend them with ways of specifying how both labels and data change. From the complexity point of view, the right type of formalisms are subclasses of register automata. These, however, are not well suited for querying. To overcome this, we develop several types of extended regular expressions to specify paths with data, and study their querying power and complexity."

- Regular path queries.
- Formal languages.