Abstract

This report gives an overview of my publications by topic. As such, it can serve as a reading guide.

Keywords: Data interoperability, data integration, probabilistic databases, object-oriented databases, information extraction, named entity recognition and disambiguation, entity resolution, data cleaning.

1. Introduction

Disclaimer: This report is not complete and should be considered work in progress. The structure indicates what you may expect where.

2. Description of Research Area “Data Interoperability”

TODO: Insert contents of data interoperability description here

3. Probabilistic data integration

Vision: JIT [10].

3.1. Probabilistic XML

Quality measures: [2].

3.2. Indeterministic deduplication

Deduplication: JDIQ [9].

3.3. Web harvesting

Size estimation of deep web sources: iiWAS [7], DB-DBD [8], TR [6].

4. Information Extraction

NER & NED with feedback loop: KDIR main [3], ICWE demo [4], SWAIE Twitter [5].

5. XML databases

6. Object-Oriented / Complex Object Databases

7. Other work

Recommender systems: [1].

TODO: Also include references to a selection of master theses

References


