

Overview and reading guide to my publications

Maurice van Keulen^{a,*}

^a*University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*

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], DBDBD ↓[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

- [1] V. de Graaff, M. van Keulen, R. A. de By, Towards geosocial recommender systems, in: 4th International Workshop on Web Intelligence & Communities (WI&C 2012), Lyon, France, ACM, New York, 2012, pp. 8:1–8:4.
- [2] A. de Keijzer, M. van Keulen, Quality measures in uncertain data management, in: H. Prade, V. S. Subrahmanian (eds.), Proceedings of the First International Conference on Scalable Uncertainty Management (SUM2007), Washington, DC, USA, vol. 4772 of Lecture Notes in Computer Science, Springer Verlag, Berlin, 2007, pp. 104–115.
- [3] M. B. Habib, M. van Keulen, Improving toponym disambiguation by iteratively enhancing certainty of extraction, in: Proceedings of the 14th International Conference on Knowledge Discovery and Information Retrieval, KDIR 2012, Barcelona, Spain, SciTePress, Spain, 2012.
- [4] M. B. Habib, M. van Keulen, Improving toponym extraction and disambiguation using feedback loop, in: Proceedings of the 12th International Conference on Web Engineering (ICWE 2012), Berlin, Germany, Lecture Notes in Computer Science, Springer Verlag, Berlin, 2012, pp. 439–443.
- [5] M. B. Habib, M. van Keulen, Unsupervised improvement of named entity extraction in short informal context using disambiguation clues, in: Proceedings of the EKAW Workshop on Semantic Web and Information Extraction (SWAIE 2012), Galway, Ireland, vol. 925 of CEUR Workshop Proceedings, CEUR-WS.org, Germany, 2012, pp. 1–10.
- [6] S. M. Khelghati, D. Hiemstra, M. van Keulen, Deep web entity monitoring, Technical Report TR-CTIT-13-02, Centre for Telematics and Information Technology, University of Twente, Enschede (December 2012).
- [7] S. M. Khelghati, D. Hiemstra, M. van Keulen, Size estimation of non-cooperative data collections, in: Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services (iiWAS2012), Bali, Indonesia, ACM, New York, 2012, pp. 239–246.

*Corresponding author

Email address: m.vankeulen@utwente.nl (Maurice van Keulen)

- [8] S. M. Khelghati, D. Hiemstra, M. van Keulen, Size estimation of non-cooperative data collections, in: The Dutch-Belgian Database Day 2012, Brussels, Belgium, Université Libre de Bruxelles, Brussels, Belgium, 2012.
- [9] F. Panse, N. Ritter, M. van Keulen, Indeterministic handling of uncertain decisions in deduplication, Journal of Data and Information Quality not available yet, accepted, awaiting publication.
- [10] M. van Keulen, Managing uncertainty: The road towards better data interoperability, IT - Information Technology 54 (3) (2012) 138-146.