

2023 年度年次報告書

基礎理論とシステム基盤技術の融合による Society 5.0 のための基盤ソフトウェアの創出

2022 年度採択研究代表者

天笠 俊之

筑波大学 計算科学研究センター
教授

検証可能なデータエコシステム

主たる共同研究者:

石川 佳治 (名古屋大学 大学院情報学研究科 教授)

小口 正人 (お茶の水女子大学 基幹研究院 教授)

鬼塚 真 (大阪大学 大学院情報科学研究科 教授)

宮崎 純 (東京工業大学 情報理工学院 教授)

森嶋 厚行 (筑波大学 図書館情報メディア系 教授)

研究成果の概要

検証可能なデータエコシステムの実現に向けて、以下の項目について研究を実施した。

1. メタデータマネージャの要素技術として、実体マッチング 1) およびテキスト中の実態識別手法を開発するとともに、分散 RDF 処理のためのデータ分割手法を開発した 2)。またスマート農業実証のためのデータセットとして、トマト画像からの果実検出手法を検討した。
2. データリネージュのためのデータモデルの研究とそのプロトタイプ実装、データリネージュ管理のためのグラフストレージとその高速化技術、データ間の関係の可視化のためのメタデータ抽出技術 3)等について研究を実施した。
3. データ検証が行える分散データベースの基本的な性能評価、および、分散データベースのメタデータ検証のためのメタデータ生成・更新処理の組み込みコマンドとしての実装を行った。
4. 統合データモデルに関する実現技術として、因果推論とリネージュに基づく統合データモデルの開発、品質を保証する近似的なデータベース問合せの処理手法、高次元データのための高速な最近傍問合せ処理技術等に関する研究を行なった。
5. データベース修復・シミュレーションの実現のための要素技術として、過去のトランザクション処理の変更による DB 修復・シミュレーション、LLM を用いたデータクリーニング・テーブル統合 4)・データ圧縮、プライバシーを保護するデータベース合成手法等に関する研究を実施した。
6. オープン環境において取得されるデータ入手、ヒューマンインザループによる信頼性向上やデータ検証プロセスの支援を行う具体的な複数手法の開発に着手し、ヒューマンインザループによる信頼できる開世界情報源の発見 5)、人・AI 協働による信頼できる開世界データ処理手法、ヒューマンコンピューテーションのためのデータ検証支援手法等に関する研究を実施した。

【代表的な原著論文情報】

- 1) John Bosco Mugeni, Steve Lynden, Toshiyuki Amagasa and Akiyoshi Matono, "AdapterEM: Pre-trained Language Model Adaptation for Generalized Entity Matching using Adapter-tuning," 27th International Database Engineering & Applications Symposium (IDEAS 2023), pp. 140-147, Heraklion, Crete, Greece, May 5-7, 2023.
- 2) Kosuke Yamasaki, Toshiyuki Amagasa, "RDF Data Partitioning for Efficient SPARQL Query Processing with Spark SQL," The 25th Int'l Conf. on Information Integration and Web Intelligence (iiWAS2023), pp. 92-106, Bali, Indonesia, Dec. 4-6, 2023.
- 3) Mingman Xu, Qiong Chang, Jun Miyazaki, How Does the System Perceive Me? --- A Transparent and Tunable Recommender System, The 34th International Conference on Database and Expert Systems Applications (DEXA 2023), pp. 33-48, 2023
- 4) Yuyang Dong, Chuan Xiao, Takuma Nozawa, Masafumi Enomoto, Masafumi Oyamada: DeepJoin: Joinable Table Discovery with Pre-trained Language Models. Proc. VLDB Endow. 16(10): 2458-2470 (2023)
- 5) Haruto Yamasaki, Masaki Matsubara, Hiroyoshi Ito, Yuta Nambu, Masahiro Kohjima, Yuki

Kurauchi, Ryuji Yamamoto, Atsuyuki Morishima. A Cluster-Aware Transfer Learning for Bayesian Optimization of Personalized Preference Models. The Eleventh AAAI Conference on Human Computation and Crowdsourcing (HCOMP2023), pp. 175-185, 2023.