

メンバー:

- チームリーダー: 松本裕治
- 特別研究員: 西田典起, 寺西裕紀, Van-Thuy PHI, Rumana Ferdous MUNNE
- テクニカルスタッフ: Shanshan LIU, 徳永なるみ
- 客員研究員: 新保仁、重藤優太郎(千葉工大), 須藤克仁, 大内啓樹 (NAIST), 古崎晃司(大阪電通大) 相澤彰子, 森純一郎(東大), 荒瀬由紀(東科大), 山田育矢(Studio Ousia), 芝原隆善(レトリバ), 山内健二
- パートタイマー研究員: 博士課程学生5名

概要・研究の意義

- 文書データ、主に専門分野の文書からの知識獲得技術の開発、および、専門分野オントロジーの自動構築を目的とする研究開発を行う
- 文書解析のための自然言語処理に関する基盤技術の研究開発を行う
- 専門分野の研究者との協働により実応用システムの研究開発を行う



研究項目:

- 専門文書からの知識獲得に関する基盤技術の開発および応用システムの研究開発を行う
- 材料分野における特性情報やプロセス情報の抽出手法の研究開発を行、材料データベースの半自動構築を実現する
- 専門用語、複雑な文、文章構造などの言語解析、高度な文書検索などの基礎研究を行う

これまでの代表的な研究成果:

- COVID-19関連論文の意味に基づく検索システムCovRelexをJST-CRESTの共同研究として構築
- PolyMinder: 材料論文からのデータ構造化AIツール群とユーザインタフェースの開発
- JST日独仏AI研究において薬害情報のアノテーションデータの構築
- 論文からの知識獲得に関する手法の研究開発
 - PDF論文の構造解析 / 図表の内部構造解析
 - 並列句解析、および、並列句を伴う専門用語表現の認識
 - 文書構造解析(解析システム、および、文書構造アノテーションデータの構築)
 - 関係情報の抽出(生成モデルによる関係認識)

今後の予定・目標:

- 汎用かつ分野適応可能な知識獲得: 専門文書からの用語認識、物性情報や合成プロセス抽出のための汎用かつ分野適応可能なツールを構築し、公開する
- 基盤的言語解析: 複雑な文構造の解析の精度向上を達成し、表面的な言語解析では抽出が難しい知識の獲得手法を実現する
- 文書解析: 文書構造解析を利用し、文書全体からの知識獲得技術を開発する
- 潜在的構造に基づく知識獲得: 分野オントロジーや課題のための構造獲得と並行した知識獲得とその利用

外部資金・共同研究:

- NIMS: ポリマーデータベースPolyInfoのための材料特性情報・プロセス情報の自動獲得(2016~)
- CREST: 科学技術論文解析(2015.10~2021.3)
- NEDO: 材料データ構造化AIツール(2019.7~2022.2)
- JST日独仏AI研究: 「医薬品安全性監視のための言語を越えた知識強化情報抽出」(2021.1~2024.3)
- 理研BRCとの生体プロセス情報抽出の共同研究
- 企業との共同研究・技術指導: 4件

主な研究発表:

- Rumana Ferdous Munne, Noriki Nishida, Shanshan Liu, Narumi Tokunaga, Yuki Yamagata, Kouji Kozaki1, Yuji Matsumoto, "Zero-Shot Entailment Learning for Ontology-Based Biomedical Annotation Without Explicit Mentions," COLING 2025, Abu Dhabi, UAE, January 2025.
- Dinh-Truong Do, Hoang-An Trieu, Van-Thuy Phi, Minh Le Nguyen, Yuji Matsumoto, "PolyMinder: A Support System for Entity Annotation and Relation Extraction in Polymer Science Documents," COLING 2025: System Demonstrations, Abu Dhabi, UAE January 2025.
- Van-Thuy Phi, Hiroki Teranishi, Yuji Matsumoto, Hiroyuki Oka, Masashi Ishii, "PolyNERE: A Novel Ontology and Corpus for Named Entity Recognition and Relation Extraction in Polymer Science Domain," LREC-COLING 2024, Torino, Italia, May 2024.
- T. Do, C. Nguyen, V. Tran, K. Satoh, Y. Matsumoto, M. Nguyen, "CovRelex-SE: Adding Semantic Information for Relation Search via Sequence Embedding," EACL: System Demonstrations, Dubrovnik, Croatia, May 2023.