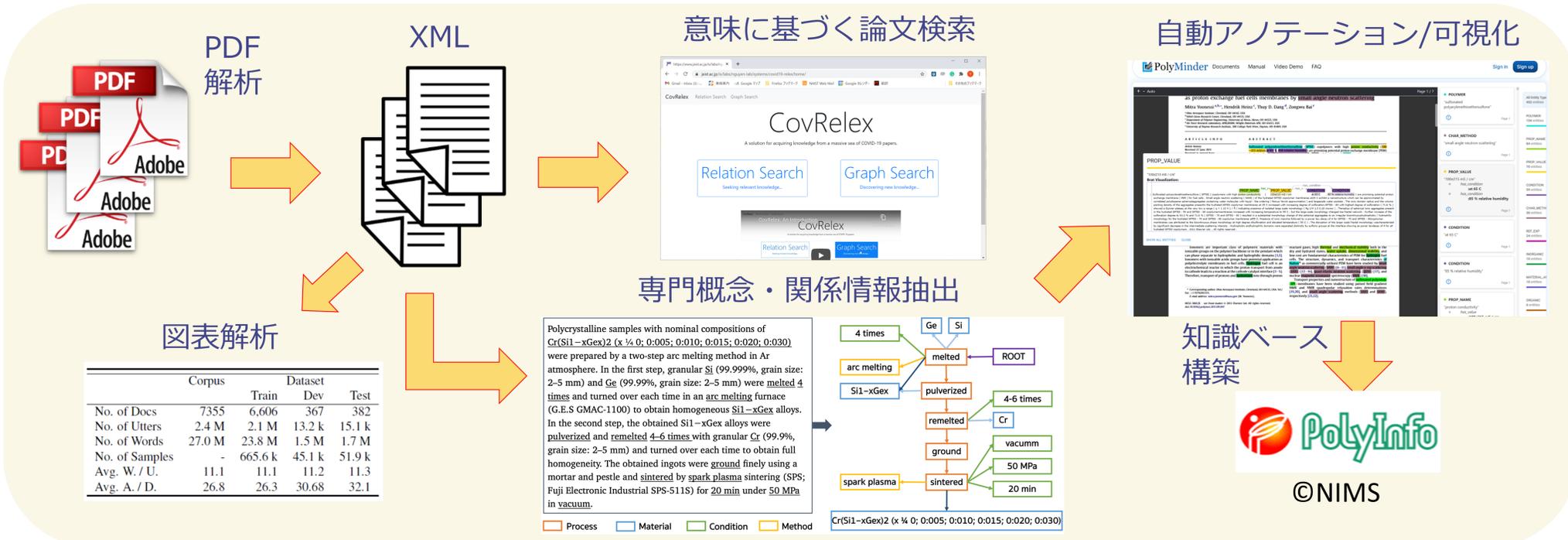


メンバー:

- チームリーダー: 松本裕治
- 研究員・特別研究員: 西田典起, 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: 11名
- 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

最近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[1]: 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(UI)の開発
- DOCORA[2]: 上記UIを他分野にも汎用化
- 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[3]
 - 専門用語認識・関係抽出システムの開発とそのための学習用アノテーションデータの構築[4,5]
 - 医学生物学分野における概念認識・関係抽出[6,7]

今後の予定・目標:

- 論文中の図表からの情報抽出手法の開発
- 分野知識を利用した知識獲得: 分野オントロジーや専門分野の知識や規則を利用した知識獲得手法の開発
- 共参照解析などテキスト理解のための基盤技術の高度化
- 学習データの利用による教師付き学習と大規模言語モデルによる教師なし学習の相補利用による実用化

関連する最近の研究発表:

- [1] 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.
- [2] Dinh-Truong Do, Hoang-An Trieu, Van-Thuy Phi, Le-Minh Nguyen, Yuji Matsumoto, "Docora: A System for Interactive Knowledge Extraction and Visualization from Scientific PDFs." AAIL-26: Demonstration Track, January 2026.
- [3] Van-Thuy Phi, Dinh-Truong Do, Hoang-An Trieu, Yuji Matsumoto, "A Hybrid LLM and Supervised Model Pipeline for Polymer Property Extraction from Tables in Scientific Literature," Proceedings of the Third Workshop for Artificial Intelligence for Scientific Publications, December 23, 2025.
- [4] Van-Thuy Pei, 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.
- [5] Van-Thuy Phi, Yuji Matsumoto, "A Unified Framework for N-ary Property Information Extraction in Materials Science," Findings of the Association for Computational Linguistics: EMNLP-2025, November 2025.
- [6] 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.
- [7] Shanshan Liu, Noriki Nishida, Rumana Ferdous Munne, Narumi Tokunaga, Yuki Yamagata, Kouji Kozaki, and Yuji Matsumoto, "MA-COIR: Leveraging Semantic Search Index and Generative Models for Ontology-Driven Biomedical Concept Recognition", ACL-2025, Student Research Workshop, July 2025.