

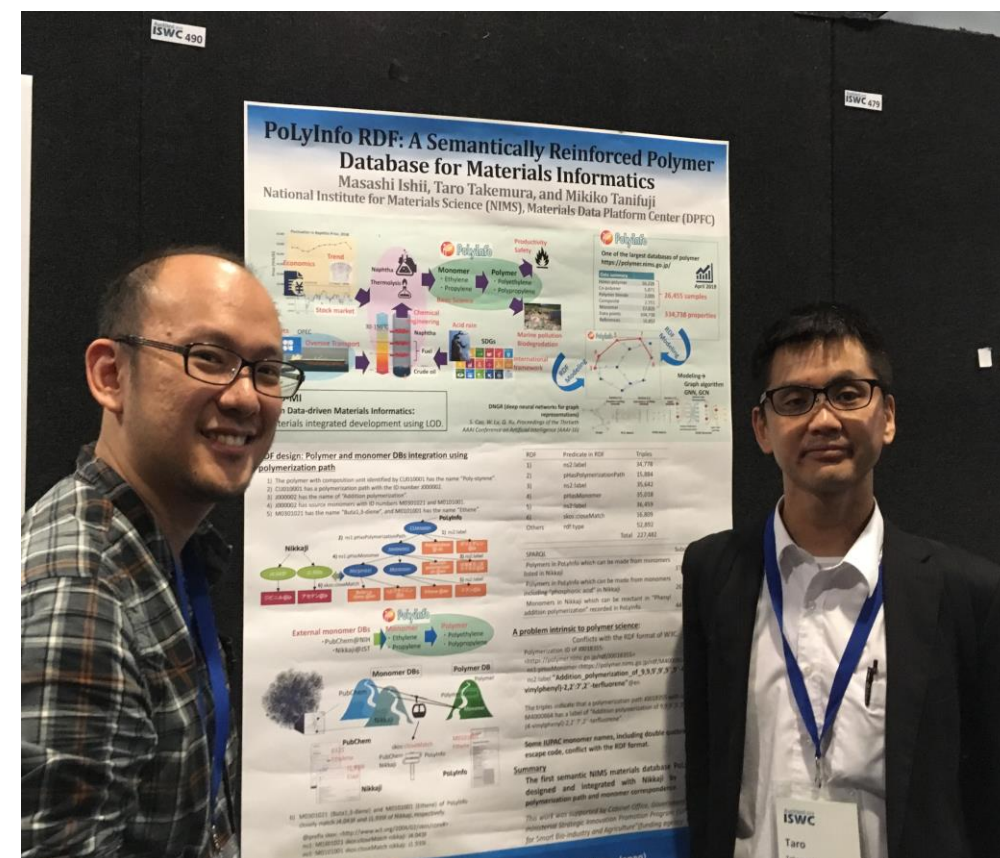
ISWC2019参加報告： データプラットフォーム構築の ウェブセマンティック技術

セッション参加者：谷藤幹子、竹村太郎、（石井真史）

物質・材料研究機構 MaDIS

材料データプラットフォームセンター

人工知能学会セマンティックウェブとオントロジー研究会ワークショップ
第一部 ISWC2019参加報告, 2019/11/22@慶応大学矢上キャンパス



■私の命題：セマンティックウェブ技術を使う知識ベース(Knowledge Base)のアプローチを材料データにあて、材料開発や物質探索、データベースの知識化に使うとしたら、何が最適で、どのような可能性があるか。

■ISWC2019の内容分類

- (1) Data Acquisition: データを抽出する
- (2) Mapping To Ontology: データからオントロジーへのマッピング (例Karma)
- (3) Data Schema format: Schema.org
- (4) Data Export format: Elasticsearch, triple を読めないbig data ツールも考慮してRDF だけでなくJSON-LD 形式も生成
- (5) Entity Linking & Similarity: エンティティリンキングでは同じ指示対象をもつエンティティを対応づけ
- (6) Knowledge Graph Deployment :ナレッジグラフを知識ベースに配置. JSON-LD 形式のドキュメントを高速全文検索・解析エンジンElasticsearch 用にインデクシング
- (7) Query & Visualization: クエリ検索と結果の視覚化でシステムのUI でターゲットとする検索情報を入力、関連情報の検索結果を表示
 - ・ SWSA Ten-Year Award
 - ・ Olaf Hartingほか: Executing SPARQL Queries over the Web of Linked Data
 - ・ Julius Voltzほか: Discovering and Maintaining Links on the Web of Data

ウィーン大学でのデータセット (タイプ) とクエリ (検索) のベンチマーク

■私の結論

持っているデータをタキソノミーで分類してみる



その分類をコンセプト化 (上位概念) でくくってみる (オントロジー)



上位からの景色で落ちた概念をオントロジーに追加する



タキソノミーに落として、データ拡張や材料データベースとRDF統合連携で、なにが・どのような粒度で・どう出てくるか評価する

■ほか：参加セッションと感想

Session 1A: Knowledge Graph Embeddings

Session 2A: Integration and Fusion

Session 3A: Deep Learning

Session 4D: Knowledge Graph Building (1)

Session 5C: Ontology Design

Session 6B: SPARQL

Session 7B: Information Extraction

Session 8A: Linked Data Analytics & Dynamics

コメント：主にKnowledge Graphを中心として、それらを繋いで統合、そこから情報や意味を抽出する技術について情報収集した。提案されていた技術手法のいくつかは材料系DBにも適用できそうである。ただし、材料系の場合はきれいなDBではなく、データの欠損やノイズ成分も多いことが課題になるように思われた。（竹）

■出張後記

本報告書は、野本氏によるISWC2015参加報告を参考に、研究の進展を理解するソースとして参考にしている。4年たった今も、同じ研究課題があり、しかし社会実装が確実に広がっていることが明白なISWC2019であった。とりわけWWWというインターネット創生の機には新興分野であった同会議が、今日のAIと高速・大容量通信インターネットによって、新しく、地球規模の知識産業にリボーン（再生）している姿が、（4000名程度の小規模ながらも）インパクトがあった。特に、技術の実用化に伴う様々なハザマ課題にも丁寧に取り組み、そして気長に（失敗も許される）研究として人材創出に取り組む欧州は、特に優れている。企業や研究プロジェクトからのポスドク採用やインターン勧誘も盛んで、実データをつかったセマンティックウェブ発展の現実の中で、研究・開発の両輪をもつNIMS DPFCビジョンは、長期的なビジョンで研究と実用化を設計する必要があるのではないか。

1. データ入力：つくる、ためる ⇔物性名等の同義語辞書、材料記述→RDF→タキソミー→オントロジー

a. データをリポジトリに登録する時のメタデータ語彙の統制

b. 登録データのメタデータ検索と表示 ⇔exact / similarity / bilingual / 密度 →セマンティックな横断・動的な処理の可能性

日本の材料データプラットフォーム

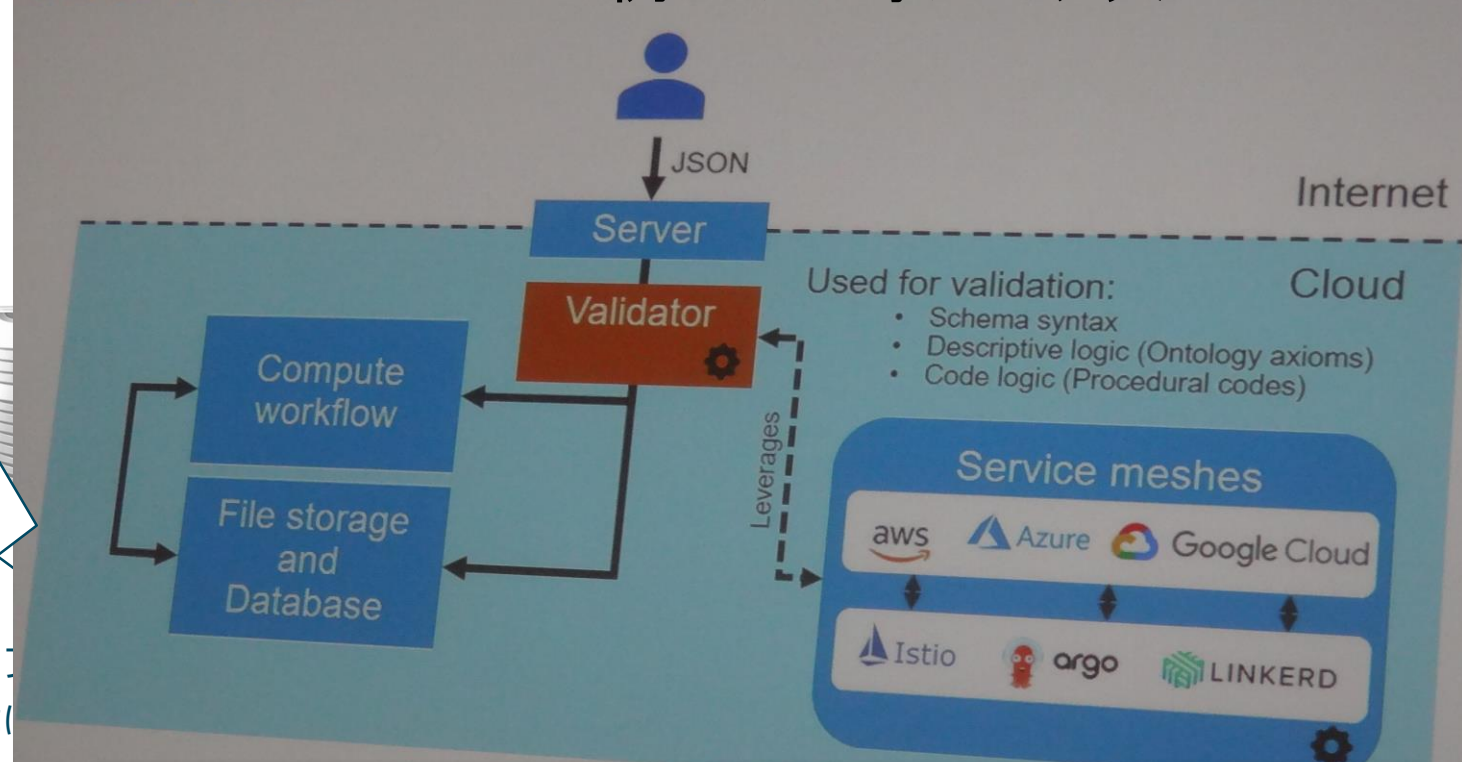
●いまやりたいこと

- 1 材料メタスキーマでデータ表現、登録データ対象に検索
- 2 メタ登録の入力支援（用語統制・選択としてWikiから）

●将来的にどうなるか

- 3 Wiki語彙をつかってアクセスできるようにしたい
- 4 RDFで統合検索、知識体系として表現したい
- 5 JSON-LODで出力、知識活用ツールに繋げたい

Data transfer 例：カルフォルニア大のFORGE



Knowledge Exploration

例：MT社のmetaphacts



- ▶ Front end framework
- ▶ Flexibilities for customization: templates, diagram, faceted search
- ▶ Graphical way to define queries

2. データ出力：材料名や合成情報など ⇔統合・融合のための動的なオントロジーの生成・チューニング検索 →GUI視認/API出力→機械学習→AIによる気づきへ