

人工知能学会第54回SWO研究会 × ナレッジグラフ若手の回  
**ISWC2020サーベイ**

2021年8月4日 オンライン

# 企画の趣旨

- ・セマンティックWebとナレッジグラフのトップカンファレンス  
ISWCの最新動向を, 有志により分担調査し, その結果を国内コミュニティで共有する
- ・ISWCへの投稿と国内研究コミュニティの活性化を図る
- ・実施方法
  - ・本会議のセッションを分担し1論文1スライドで紹介
  - ・1編の論文のサーベイ内容はアブストラクト, 序論 (+α) レベルとし, 聴衆が, 「どの論文を読めば良いか?」を決める際の参考情報とすることを想定
  - ・対象トラックはResearch, In-Use, Resources, Industry
  - ・スライドは研究会Webページで公開する

# ISWC2020情報源

- プログラム
  - <https://iswc2020.semanticweb.org/program/conference/>
- 論文
  - ISWC 2020, Part I, LNCS 12506: <https://link.springer.com/book/10.1007/978-3-030-62419-4>
  - ISWC 2020, Part II, LNCS 12507: <https://link.springer.com/book/10.1007/978-3-030-62466-8>
  - Posters, Demos and Industry Track: <http://ceur-ws.org/Vol-2721/>
- 受賞情報
  - [https://twitter.com/iswc\\_conf/status/1324763096272015360](https://twitter.com/iswc_conf/status/1324763096272015360)
  - [https://twitter.com/iswc\\_conf/status/1324763671873114120](https://twitter.com/iswc_conf/status/1324763671873114120)
  - [https://twitter.com/iswc\\_conf/status/1324764451103494145](https://twitter.com/iswc_conf/status/1324764451103494145)
  - [https://twitter.com/iswc\\_conf/status/1324764696436711424](https://twitter.com/iswc_conf/status/1324764696436711424)
  - [https://twitter.com/iswc\\_conf/status/1324765629795180545](https://twitter.com/iswc_conf/status/1324765629795180545)
  - [https://twitter.com/iswc\\_conf/status/1324766153366011908](https://twitter.com/iswc_conf/status/1324766153366011908)
  - <https://iswc2020.semanticweb.org/program/awards/>

# 担当者一覧

発表順	担当者(申込情報のまま)	担当セッション
1	江上周作 (産業技術総合研究所)	SESSION 3A: INFORMATION RETRIEVAL AND SEMANTIC ANALYSIS SESSION 3B: KNOWLEDGE GRAPHS SESSION 3C: KNOWLEDGE ENHANCED LEARNING
2	野本昌子 (理化学研究所)	SESSION 1C: INFORMATION RETRIEVAL AND SEMANTIC ANALYSIS SESSION 4A: INFORMATION RETRIEVAL AND SEMANTIC ANALYSIS
3	カワムライクエ (名古屋工業大学)	SESSION 9C: DATA MINING AND KNOWLEDGE-DRIVEN ML
4	藤原浩司 (株式会社東芝)	SESSION 5A: ONTOLOGY-BASED INTEGRATION AND EXCHANGE SESSION 7D: ROBUST AND SCALABLE MANAGEMENT OF SEMANTICS AND DATA
5	砂川英一 (株式会社東芝)	SESSION 5D: KNOWLEDGE GRAPHS IN USE SESSION 6B: KNOWLEDGE DRIVEN NLP
6	勝島修平 (東京都市大学大学院 総合理工学研究科情報専攻)	SESSION 1B: KNOWLEDGE GRAPHS SESSION 4B: DATA MINING AND KNOWLEDGE-DRIVEN ML
7	鵜飼孝典 (富士通)	SESSION 9B: MACHINE LEARNING FOR THE SEMANTIC WEB SESSION 6C: ONTOLOGY ENGINEERING
8	隅田飛鳥 (AIP)	SESSION 2A: SHACL
9	小柳祐介 (富士通株式会社)	SESSION 3D: SEARCH, QUERY, INTEGRATION AND ANALYSIS SESSION 2C: DATA MINING AND KNOWLEDGE-DRIVEN ML
10	亀田堯宙 (国立歴史民俗博物館)	SESSION 10A: OPEN DATA SESSION 1D: SCHOLARLY DATA
11	浅野優 (日立製作所)	SESSION 7B: INDUSTRY 2 SESSION 10B: QUESTION ANSWERING
12	長野伸一 (東芝)	SESSION 6D: INDUSTRY 1 SESSION 10D: IN-USE APPLICATIONS

# 残りセッション

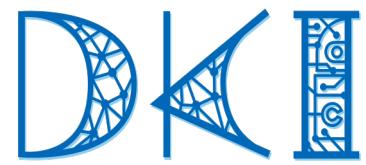
- SESSION 1A: KNOWLEDGE REPRESENTATION AND REASONING
- SESSION 2B: KNOWLEDGE GRAPHS
- SESSION 2D: PROGRAMMING THE SEMANTIC WEB
- SESSION 4C: QUERY, INTEGRATION AND ANALYSIS
- SESSION 4D: DATA STREAMS
- SESSION 5B: REASONING IN USE
- SESSION 5C: DATA MINING AND KNOWLEDGE-DRIVEN ML
- SESSION 7C: ONTOLOGY ENGINEERING
- SESSION 9D: ONTOLOGY MAPPING, MERGING, AND ALIGNMENT
- SESSION 10C: INDUSTRY 3

SESSION 3A: INFORMATION RETRIEVAL AND  
SEMANTIC ANALYSIS

SESSION 3B: KNOWLEDGE GRAPHS

SESSION 3C: KNOWLEDGE ENHANCED LEARNING

1. 江上周作（産業技術総合研究所）



# ISWC2020 サーベイ

**SESSION 3A: INFORMATION RETRIEVAL AND SEMANTIC ANALYSIS**

**SESSION 3B: KNOWLEDGE GRAPHS**

**SESSION 3C: KNOWLEDGE ENHANCED LEARNING**

産業技術総合研究所  
江上周作

# Google Dataset Search by the Numbers



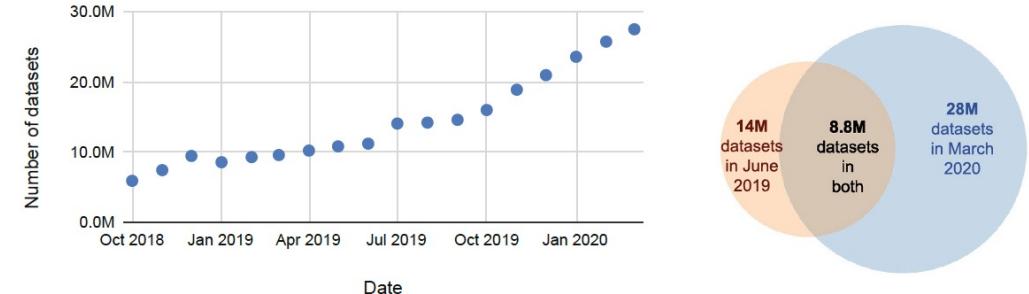
- Authors: Omar Benjelloun, Shiyu Chen and Natasha Noy (Google Research, USA)

## 概要

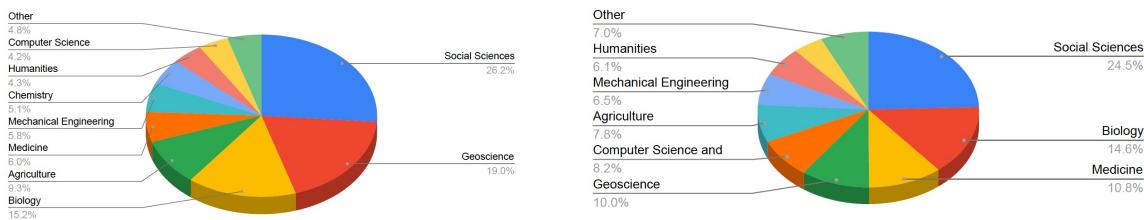
- Google Dataset Searchを2016年に開始してから schema.orgで記述されたデータセットの数は 500K→30Mに増加
- Dataset SearchはWeb上の意味づけされたデータセットのメタデータコレクションとして最大規模

## 貢献

- 2800万個のデータセットについてメタデータのコーパスを分析する手法を示す
- このようなコーパスが分析に役立つResearch questionsを特定し、コーパスの分析結果を示す
- コーパス分析からのlesson learnedを議論



Category	Number of datasets	% of total	Sample formats
Tables	7,822K	37%	CSV, XLS
Structured	6,312K	30%	JSON, XML, OWL, RDF
Documents	2,277K	11%	PDF, DOC, HTML
Images	1,027K	5%	JPEG, PNG, TIFF
Archives	659K	3%	ZIP, TAR, RAR
Text	623K	3%	TXT, ASCII
Geospatial	376K	2%	SHP, GEOJSON, KML
Computational biology	110K	<1%	SBML, BIOPAX2, SBGN
Audio	27K	<1%	WAV, MP3, OGG
Video	9K	<1%	AVI, MPG
Presentations	7K	<1%	PPTX
Medical imaging	4K	<1%	NII, DCM
Other categories	2,245K	11%	



# Facilitating COVID-19 Meta-analysis Through a Literature Knowledge Graph



- Authors: Bram Steenwinckel, Gilles Vandewiele, Ilja Rausch, Pieter Heyvaert, Ruben Taelman, Pieter Colpaert, Pieter Simoens, Anastasia Dimou, Filip De Turck and Femke Ongenae (Ghent University, Belgium)

## 概要

- 新型コロナのパンデミックに対応するため、63,000件を超える論文の情報を含むデータセットがKaggleで公開された <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>
- 目的はこのデータセットをKG化し、メリットを享受すること
  - 外部リソースとつながることで知識を強化。エッジはエンティティ間関係を明示的に表すため、より貴重な洞察をもたらす。

## 貢献

- KG構築のフルパイプラインを示し、FAIR (Findable, Accessible, Interoperable, and Reusable)原則を考慮してリソースの公開化方法を説明
- 予備的な分析を行ってこのKGの利点を説明

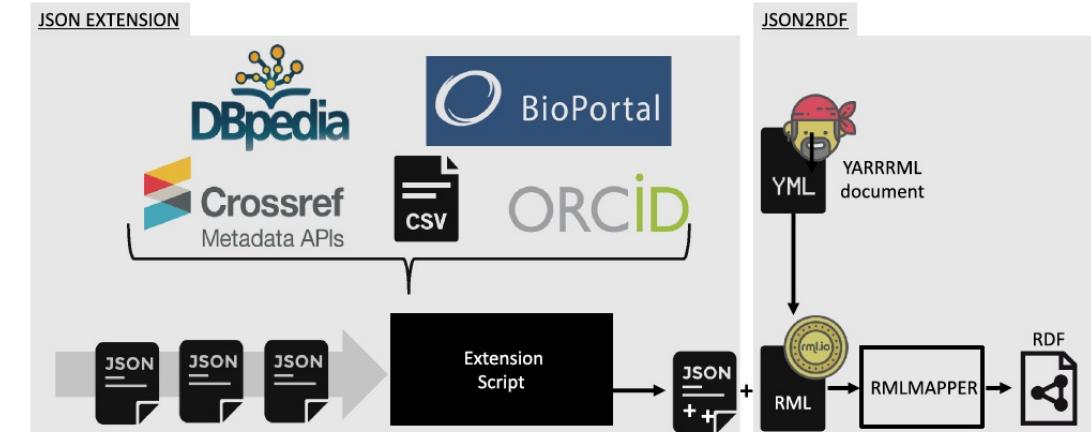


Fig. 2. Overview of each component used to transform the dataset from Sect. 3.1 into an RDF representation.

# The Virtual Knowledge Graph System Ontop



- Authors: Guohui Xiao, Davide Lanti, Roman Kontchakov, Sarah Komla Ebri, Elem Guzel Kalayci, Linfang Ding, Julien Corman, Benjamin Cogrel, Diego Calvanese and Elena Botoeva (Free Univ. of Bozen-Bolzano, Italy., Univ. of London, UK., Virtual Vehicle Research GmbH, Austria., Umeå Univ. Sweeden., Imperial College London, UK.)

## • 概要

- Virtual Knowledge Graph (VKG)のアプローチはOntology-Based Data Access (OBDA)として知られている
- リレーションナルデータベースをマッピングやオントロジーにより統合されたKGとして仮想化することで、エンドユーザは使い慣れた語彙でクエリを実行可能

## • 貢献

- 異種データソースを統合されたKGとして公開するためのオープンソースのVKGシステム Ontop v4を公開
  - ほぼすべてのSPARQL1.1, R2RML, OWL2QL, SPARQL含意レジーム, SPARQL1.1 HTTP Protocolをサポート
- Ontop v4の設計、評価、コミュニティの取り組み、採用について説明

<https://github.com/ontop/ontop>

# KGTK: A Toolkit for Large Knowledge Graph Manipulation and Analysis



- Authors: Filip Ilievski, Daniel Garijo, Hans Chalupsky, Naren Teja, Yixiang Yao, Craig Rogers, Ronpeng Li, Jun Liu, Amandeep Singh, Daniel Schawbe and Pedro Szekely (Univ. of Southern California, USA., Pontificia Universidade Católica Rio de Janeiro, Brazil)

## • 概要

- KGは主流技術となったが、大規模にKGを操作するためのRDF, SPARQL中心のツールセットは様々であり、統合が難しい。
  - データベース: RDF triple store, Neo4J等、操作: graphy, RDFlib等、エンティティリンク: WAT, BLINK等、エンティティ解決: MinHash-LSH, MFIBlocks等、埋め込み: PyTorch-BigGraph等、グラフ分析: graph-tool, NetworkX等
  - 大規模KGを操作、検証、分析するためのフレームワークKnowledge Graph Toolkit (KGTK)を開発

## • 貢献

<https://github.com/usc-isi-i2/kgtk/>

- KGをハイパーグラフとして表すことができるKGTKファイル形式の開発
- インポートモジュール : N-Triples, Wikidata修飾用語, ConceptNet等異なるフォーマットをKGTK形式に変換
- グラフ操作モジュール : KGのvalidate, clean, filter, join, sort, mergeをする
- クエリと分析モジュール : 中心性、連結成分の計算、RoBERTa, BERT, DistilBERTを使用してテキストベースのグラフ埋め込み
- エクスポートモジュール : KGTKフォーマットをRDF、Neo4J用property graph、GMLなどに変換
- Unixパイプに基づいて複数のKG操作を構成するためのフレームワーク

# Ontology-Enhanced Machine Learning: a Bosch Use Case of Welding Quality Monitoring



- Authors: Yulia Svetashova, Baifan Zhou, Tim Pychynski, Stefan Schmid, York Sure-Vetter, Ralf Mikut and Evgeny Kharlamov (Bosch Corporate Research, Germany., Bosch Center for Artificial Intelligence, Germany)

## 概要

- 自動車産業では溶接の品質モニタリングが重要であり, Boschではそのための機械学習アプローチを開発するため、右図のワークフローを採用
  - 複雑でコストがかかり 3 つの課題:
    - (C1)コミュニケーション、(C2)データ統合、(C3)ML品質モデルの汎用性

## 貢献

- 上記課題解決のため、従来のMLワークフローをセマンティックコンポーネント（図1のグレーボックス部分）で拡張したSemMLと呼ばれるシステムを開発
  - Ontology extender: テンプレートを埋めることで専門家に上位オントロジーの観点でドメインを記述できるようにする
  - Domain knowledge annotator: アノテーションにより生データをドメインオントロジーにマッピングすることでデータ統合
  - Machine learning annotator: 自動推論によりマッピングからML関連情報の推論、各生データソースごとにMLオントロジーとデータ間のマッピングを作成
  - Ontology interpreter: MLモデルと生データの統一化された説明可能な検査を可能にする

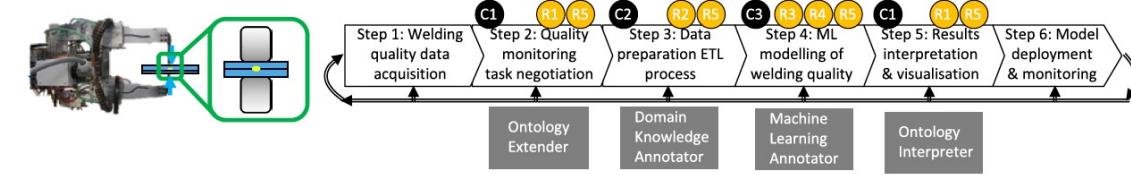


Fig. 1. A machine for automated welding (left) and an ML workflow enhanced with our semantic modules for welding quality monitoring (right). C stands for *challenges*, and R for *requirements*, see Sect. 1. ETL stands for *Extract, Transform, Load*. (Color figure online)

# NEO: A Tool for Taxonomy Enrichment with New Emerging Occupations



- Authors: Anna Giabelli, Lorenzo Malandri, Fabio Mercurio, Mario Mezzanzanica and Andrea Seveso (Univ. of Milano-Bicocca, Italy)

## 概要

- オンライン求人 (OJV) から新しい職業やスキルの需要の変化を理解できるため関心が集まる
- OJVからの知識抽出はタクソノミーの存在に大きく依存し、MLやNLPタスクに必要不可欠
  - 主要リソース: European Skills, Competences, Qualifications and Occupations (ESCO)
- タクソノミーを手動で更新するには時間がかかりエラーも発生しやすい

## 貢献

- ESCOタクソノミーを充実させるためのシステムNEOを開発
  - ドメインに依存しない指標としてHierarchical Semantic Relatedness (HSR)を定義
  - タクソノミーとOJVの両方の語彙をエンコードするためのベクトル空間モデルを合成・評価し、新たに出現する可能性のある職業の抽出、異なる分類概念のエンティティとしてのそれらの適合性を推定するための尺度GASC (Generality, Adequacy, Specificity, Comparability)を定義
  - ユーザに、提案された言及を投票するためのWebツールを提供し、タクソノミーの拡張活動において専門家を支援し、スキルギャップ分析を通じて提案された新しい職業の根拠を説明する

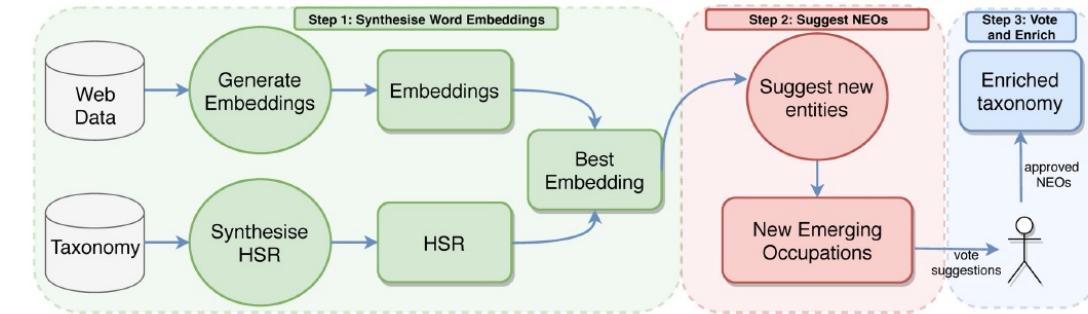


Fig. 2. A representation of the NEO workflow highlighting the main modules.

# Understanding Data Centers from Logs: Leveraging External Knowledge for Distant Supervision



- Authors: Chad DeLuca, Anna Lisa Gentile, Petar Ristoski and Steve Welch (IBM Research Almaden, USA)

## 概要

- 優れたドキュメント化が行われていない状況では、データセンター内で、どのソフトウェアプロセスがどこで実行しているかを理解することや、データセンターの移行は困難な課題
- 接続や依存関係を発見して理解するには非常に手間がかかるが、従来のツールではこの計画段階をサポートするものがほぼ無い

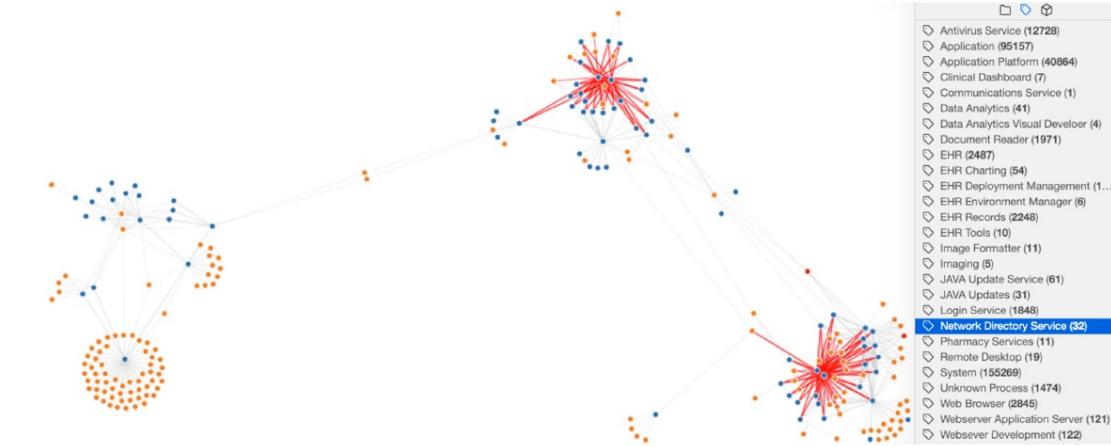


Fig. 3. Data Center Explorer - example view of a portion of the datacenter graph.

## 貢献

- データセンターの意味表現を作成するためのHuman-in-the-loopモデル知識抽出手法の提案
  - LODクラウドからソフトウェアプロセスに関する知識を収集し、
  - それをDistant supervision方式で使用してデータセンター内の各ノードに初期タグを生成、
  - 専門家が検証し、カバレッジが満足行くものになるまで繰り返す

SESSION 1C: INFORMATION RETRIEVAL AND  
SEMANTIC ANALYSIS

SESSION 4A: INFORMATION RETRIEVAL AND  
SEMANTIC ANALYSIS

2. 野本昌子 (理化学研究所)

# **ISWC2020 Survey**

SESSION [1C][4A]: INFORMATION RETRIEVAL AND SEMANTIC ANALYSIS

---

MASAKO NOMOTO

2021.8.4

RIKEN

**Session 1C:**  
**INFORMATION RETRIEVAL AND**  
**SEMANTIC ANALYSIS**

---

# [1C] PreFace: Faceted Retrieval of Prerequisites Using Domain-Specific Knowledge Bases

[Research] Prajna Upadhyay and Maya Ramanath

keywords: Facets, Prerequisite, Knowledge base

## issue

- Prior arts on prerequisite identification do not consider the multiple facets\* of the concepts.
- Existing facet generation techniques use open-domain KB, in which domain-specific relationships are under-represented.

## idea

- Extract facets together with prerequisites using a domain-specific KB and a corpus of research papers

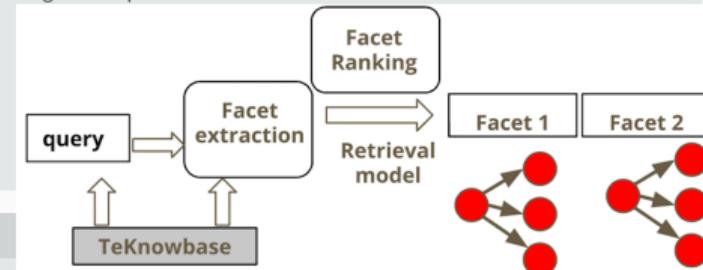
## contribution

- Introduction of the novel problem of faceted retrieval of prerequisites
- Development of the language model framework to retrieves facets and prerequisites for a concept query using a domain-specific KB and corpus and ranks them balancing the relevance and diversity.
- Demonstrating that the language model framework extracts better facets and prerequisites than using SOTA techniques separately.

## future work (in presentation)

- Generation of lecture notes for a query, summarization of facets

Fig 1. Components of PreFace



\*Facet: set of semantically similar prerequisites for the query and the aspect (presentation file p.13)

# [1C] Focused Query Expansion with Entity Cores for Patient-Centric Health Search

[Research] <spotlight paper> Erisa Terolli, Patrick Ernst and Gerhard Weikum

keywords: Query expansion, Knowledge graph, Prize Collecting Steiner Tree problem

## issue

- In health forums, expressive querying for patient-centric information needs is poorly supported by search engines.
  - ▶ IR approach (query expansion) will lead to broad and noisy queries.
  - ▶ ML approaches would not work for highly individualized needs, for the training data will be scarce and noisy.
  - ▶ Semantic-Web approaches may identify named entities and provide a crisp entity aware query using KG, but the query may not keep its focus, for user posts often contain cues for remotely related entities.

## idea

- Focused expansion using entity cores: map the task of refocusing the expanded query into a Prize Collecting Steiner Tree problem to extract the most informative and focused sub-graph from the query graph

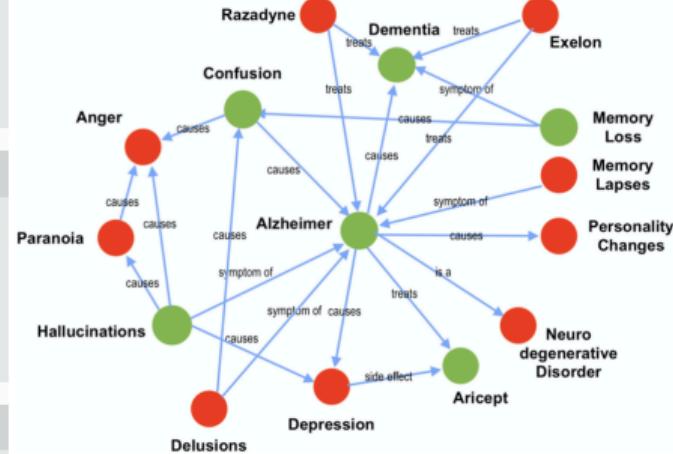
## contribution

- A new query expansion method, by computing entity cores that identify the most relevant and coherent terms for focused expansion
- Demonstrating the superiority of the method to baselines of entity-aware query expansion in search over health forums and clinical trials.

## future work

- Adapt the method to other types of health documents and other domains

Fig 4. Query Graph with Entity-Core Nodes Depicted in Green Color



# [1C] Generating Referring Expressions from RDF Knowledge Graphs for Data Linking

[Research] Armita Khajeh Nassiri, Nathalie Pernelle, Fatiha Saïs and Gianluca Quercini

keywords: Knowledge graphs, Referring expressions, Data linking

## issue

- Most of existing approaches to discover referring expressions (REs) neither scale to large KGs nor are suited for the data linking task.

## idea

- to reduce the search space, define types of graph patterns and quality measures focused on REs that are more suitable in a data linking task
- find REs from maximal non-key <sup>†</sup> set of a class

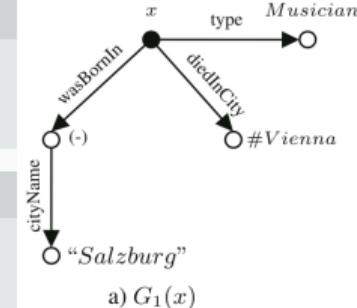
## contribution

- Defining graph patterns and several quality criteria that set forth REs, potentially relevant for data linking.
- Proposing an algorithm to discover REs for all instances of a class in a KG. The generated REs are adapted to a data linking task through the notions of minimality and diversification and the post-processing step of expansion.
- Experimental results showing that the approach can significantly increase the recall.

## future work

- Refine REs by virtue of data linking to be used for detecting missing information in the source data

Fig 1. Two graph patterns. a)



<sup>†</sup>A maximal non-key for a class C in a KG G is a set of properties P such that P is not a key, but the addition of any property to P makes it a key for that class.

**Session 4A:**  
**INFORMATION RETRIEVAL AND**  
**SEMANTIC ANALYSIS**

---

## [4A] Generating Compact and Relaxable Answers to Keyword Queries over Knowledge Graphs

[Research] Gong Cheng, Shuxin Li, Ke Zhang and Chengkai Li  
keywords: Keyword search, Knowledge graph, Query relaxation

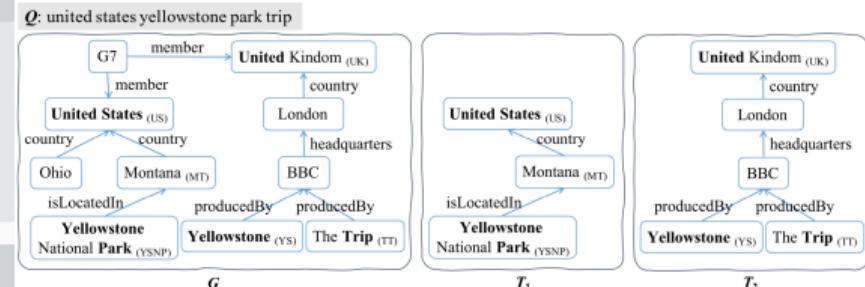
### issue

- Though users prefer structurally compact answers in exploratory search tasks, most existing methods extract answer subgraphs of group Steiner trees, which may have large graph structures.

### idea

- Computing compact and relaxable subgraphs  
(MRA: minimally relaxed answer)
  - Guaranteed answer compactness: having a bounded diameter
  - Maximized answer completeness: covering the largest number of query keywords

Fig 1. A keyword query  $Q$  over a KG  $G$  with two answers  $T_1$  and  $T_2$



### contribution

- Formulating a new combinatorial optimization problem of computing a MRA with a compactness guarantee to generate compact and relaxable subgraphs as answers
- Designing a best-first search algorithm to solve MRA
- Demonstrating the necessity of trading off answer completeness for compactness with public KGs and keyword queries

### future work

- incorporate vertex and edge weights to exploit the semantics of different types of entities and relations

# [4A] KnowlyBERT - Hybrid Query Answering over Language Models and Knowledge Graphs

[Research]: Jan-Christoph Kalo, Leandra Fichtel, Philipp Ehler and Wolf-Tilo Balke

keywords: Query answering, Language models, Knowledge graphs

## issue

- Modern knowledge graphs (KGs) lack a good coverage for most relations.
- Standard techniques for dealing with incomplete KGs:
  - ▶ (a) Relation extraction (RE) need massive amounts of training data and cannot be directly used in an on-the-fly QA system.
  - ▶ (b) KG embeddings have problems to succeed beyond baseline datasets.

Fig.1. An overview of the query answering system KnowlyBERT.

## idea

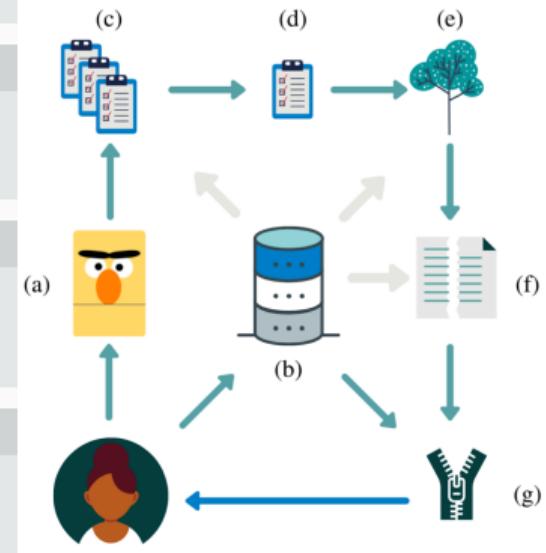
- A hybrid query answering system integrating the advantages of KGs and masked language models to cope with the incompleteness of the real-world KGs

## contribution

- Development of the hybrid system to answer entity-centric SPARQL queries.
- The experimental results show higher precision than SOTA techniques.

## future work (in the presentation)

- Which knowledge is covered by language models?
- How do we pose complex SPARQL queries to language models?, etc.
- Is there other ways how language models may support knowledge graphs?



## [4A] Linked Credibility Reviews for Explainable Misinformation Detection

[Research] <spotlight paper> Ronald Denaux and Jose Manuel Gomez-Perez

**Keywords:** Disinformation detection, Credibility signals, Explainability, Composable semantics

## issue

- Misinformation detection systems to help the general public to assess the credibility of web contents are required to output explainable results.
  - Most proposed systems, based on DL and fine-tuned to specific domains, are difficult to interpret and produce results which are not machine readable.

idea

- Linked data model for composable and explainable misinformation detection based on Credibility Reviews<sup>‡</sup> (CRs) networks of distributed bots that collaborate for misinformation detection and build a graph of CRs for web content.

## contribution

- A data model and architecture of distributed agents for composable CRs, incl. an extension to schema.org to support provenance and explainability
  - Evaluation results establishing SOTA without using the training data

## future work

- Crowdsourcing-based evaluation of the generated explanations, etc.

Fig.2 Depiction of aced bots collaborating to produce a CR for a tweet.

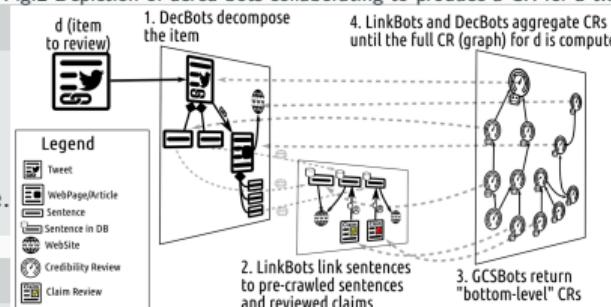
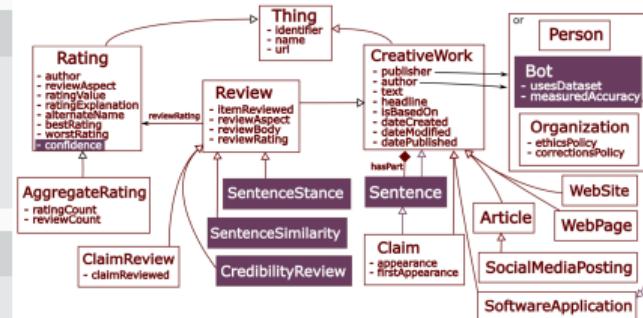


Fig.1 Linked credibility review data model, extending schema.org.



<sup>‡</sup>a tuple of (d) data item, (c) credibility rating, (r) rating confidence, (p) provenance info.

# SESSION 9C: DATA MINING AND KNOWLEDGE-DRIVEN ML

3.カワムライクエ（名古屋工業大学）

## From Syntactic Structure to Semantic Relationship: Hypernym Extraction from Definitions by Recurrent Neural Networks Using the Part of Speech Information

Tao Jia (Southwest University in Chongqing)

- hypernym（下位語）とhypernym（上位概念）の関係は、semantic networkにおいて必須の要素、定義からhypernymを特定することは、自然言語処理やセマンティック分析において重要なタスク。本研究は、ハイパニム抽出のための有用なツールを提供するだけでなく、意味的関係を学習するために構文構造を利用する例を示す。
- **関連研究**：WordNetのような公共の辞書は一般的な単語には有効だが、ドメイン固有での適用には限界がある、上位概念抽出の既存のツールにも限界がある。
- **提案手法**：単語の品詞から定義された構文構造と、学習カーネルとして双方向ゲートディリカレントユニットネットワークの両方を組み合わせた手法を提案（構文特徴の表現、構文特徴の学習、ハイパニムの識別の3つから構成）。統語的特徴をより学習するため標準的なツールを用いて各単語のPoSをラベル付けした後、定義文を品詞（PoS）列に移す。

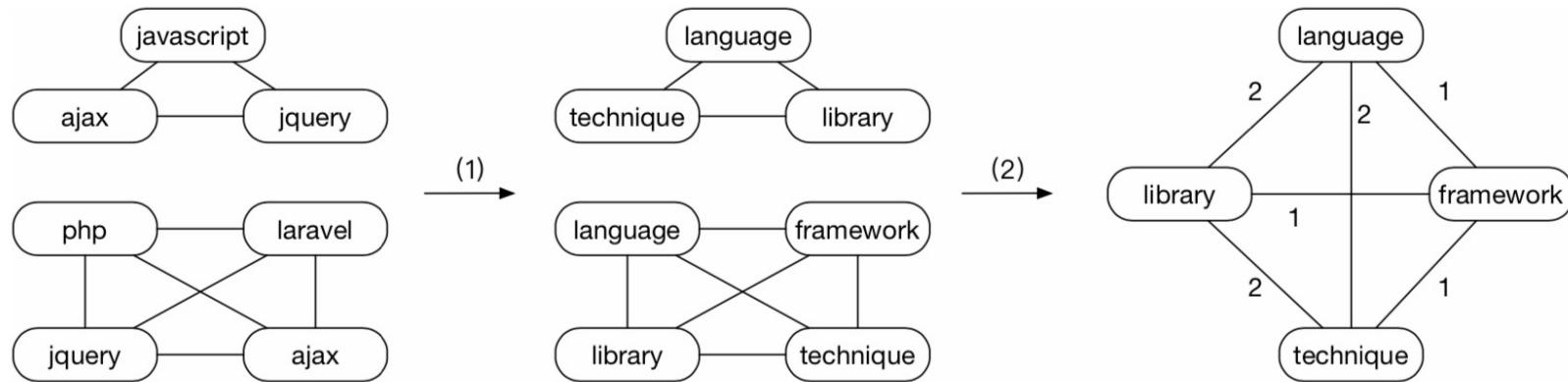


Fig.3:hypernym graph construction processの簡単な例。(1)：共起語の用語が、学習セットからの対応するhypernymに置き換えられる。(2): hypernymの共起に基づいて、hypernym共起ネットワークが構築される。

# Learning Short-term Differences and Long-term Dependencies for Entity Alignment

Jia Chen, Zhixu Li, Pengpeng Zhao, An Liu, Lei Zhao, Zhigang Chen, and Xiangliang Zhang  
 (School of Computer Science and Technology, Soochow University, China)

- KGの短期的な差異と長期的な依存性の両方を捉えるための新しいアプローチを提案
- 関連研究：KGの埋め込みベース技術は以下の3つのカテゴリに分類される

**TransE-based Entity Alignment**：最も代表的なモデル。トリプルを個別に扱う既存のTransEベースの手法の活用

**GNN-based Entity Alignment**：エンティティをグラフノードと見なしGNNを用いたエンティティアライメントを提案。

**Sequence-based Entity Alignment**: RSNは関係の出力隠れ状態に配列中の直接の前のエンティティからの残差を学習

- 提案手法：問題の定式化と、長期的な依存関係の把握（具体的には、node2vec[7]のアイデアを利用し、深さバイアスと度数バイアスを導入）および複合的損失（エンティティの長期的な情報と短期的な複合情報の両方を保持する）を行う

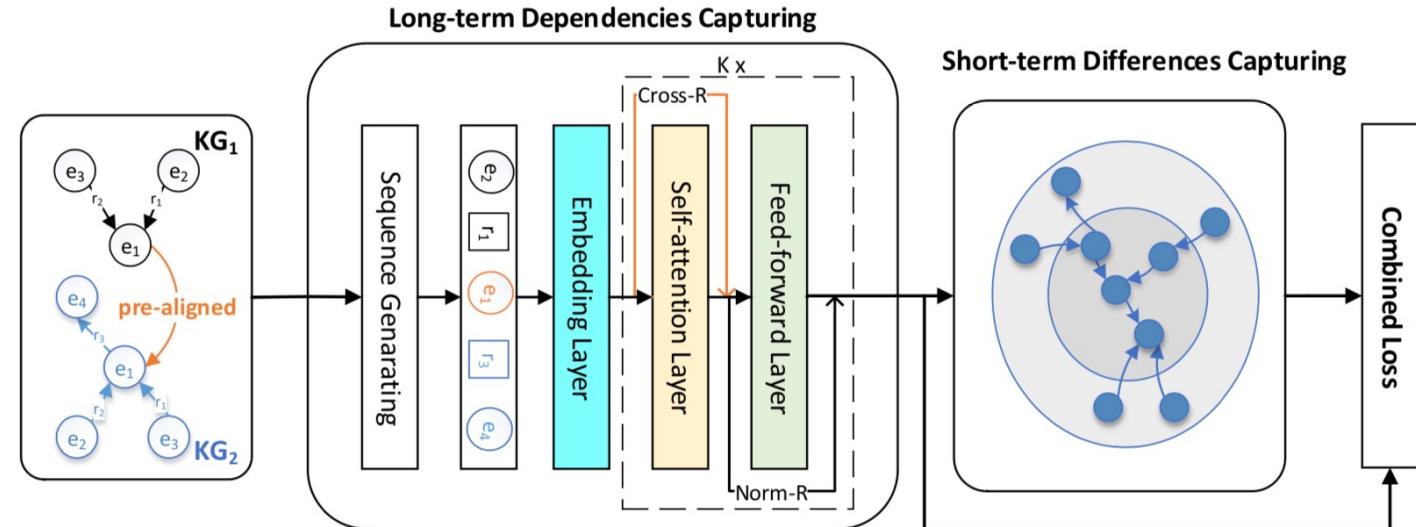


Fig. 2. The architecture of our proposed model

# Temporal Knowledge Graph Embedding Model based on Additive Time Series Decomposition

Chenjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Yazdi, and Jens Lehmann  
 (Smart Data Analytics Group, University of Bonn, Germany)

- ほとんどのKG埋め込みモデルは、時間を意識しないトリプルから学習するが、トリプル以外にも時間的な情報を含めることで、KGEモデルの性能をさらに向上させることができる。そのため、Additive Time Series decompositionを用いて、時間情報をエンティティ/リレーション表現に組み込む、時間的KG埋め込みモデルであるATiSE3を提案する。
- 関連研究**：最先端のKGEモデルの例としては、TransE、TransH、TransCom-plEx、RotatE、DistMult、ComplEx、ComplEx-N3、QuatEなどがある。これらの手法は、KGのリンク予測では良好な結果を得ているが、時間を意識していないためTKGEの推論に限界があります。
- 提案手法**：ATiSEは、トリプルのエンティティ間の関係性を利用するだけでなく、付加的時系列分解を用いて関連する時間的なメタデータを組み込む

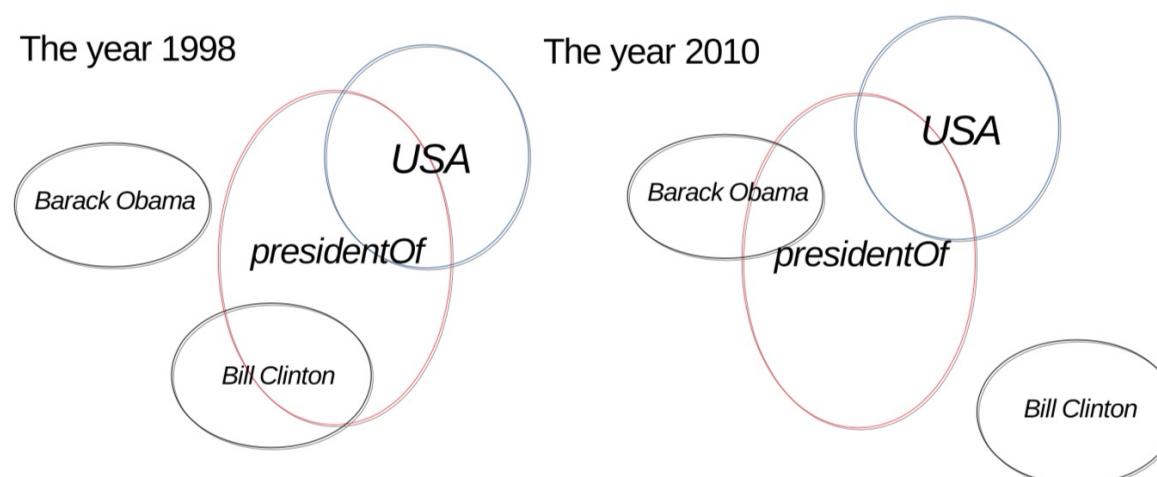


Fig.1 時間的なガウス埋め込み空間におけるエンティティと関係の平均と（対角線上の）分散の図。

SESSION 5A: ONTOLOGY-BASED  
INTEGRATION AND EXCHANGE  
SESSION 7D: ROBUST AND SCALABLE  
MANAGEMENT OF SEMANTICS AND DATA

藤原浩司 (株式会社東芝)

# ISWC2020サーベイ

SESSION 5A: ONTOLOGY-BASED INTEGRATION AND  
EXCHANGE

SESSION 7D: ROBUST AND SCALABLE MANAGEMENT OF  
SEMANTICS AND DATA

藤原浩司(東芝)

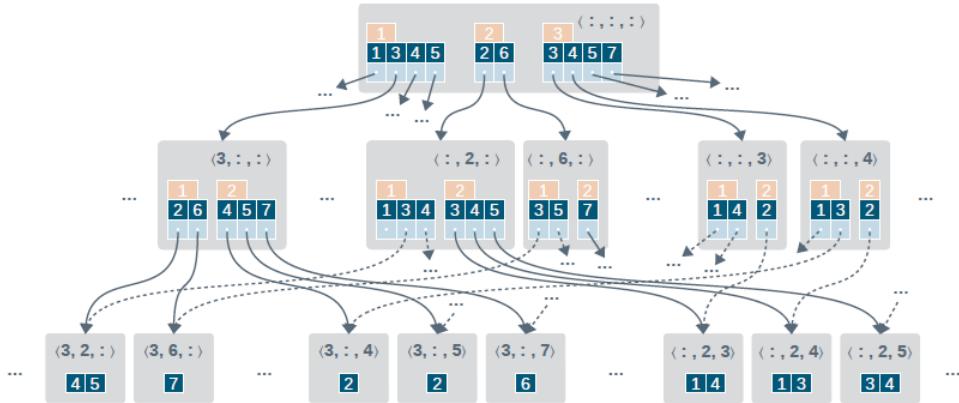
# **SESSION 5A:**

# **ONTOLOGY-BASED INTEGRATION AND**

# **EXCHANGE**

# Tentris – A Tensor-Based Triple Store(spotlight paper)

- 概要
  - テンソルベースのRDFストア Tentrisの提案
- 目的
  - 巨大なRDFに対する効率的な蓄積と保存をしたい
- 手法
  - RDFグラフをスパースな3階テンソルで表現
  - テンソル代数を使ってSPARQLを実行
- 結果
  - 10億トリプル以上の3種類のデータセットを使った実験で他の5種類のRDFストアよりも1秒あたりのクエリ解決数が最低でも1.8倍を記録した。



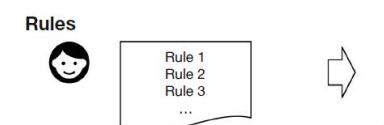
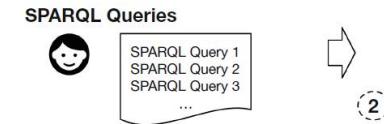
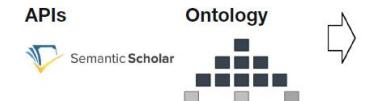
# Tab2Know: Building a Knowledge Base from Scientific Tables

- 概要
  - 論文の表から知識ベースを構築するTab2Knowの提案
- 目的
  - 研究者が論文から知見を得ることを助ける
- 手法
  - 弱教師あり学習によるテーブルの解析
  - 推論による異なるテーブルのエンティティをリンク
- 結果
  - 65%の有用なエンティティを抽出(サンプリングしたデータに対して手動で評価)
  - 97%の有用なエンティティのリンクを生成

TABLE I. RANKING OF SUBMITTED METHODS TO TASK 1.1

Method Name	Recall (%)	Precision (%)	F-score
USTB_TexStar	<b>82.38</b>	<b>93.83</b>	<b>87.74</b>
TH-TextLoc	75.85	86.82	80.96
I2R_NUS_FAR	71.42	84.17	77.27
<i>Baseline</i>	69.21	84.94	76.27
Text Detection [15], [16]	73.18	78.62	75.81
I2R_NUS	67.52	85.19	75.34
BDTD_CASIA	67.05	78.98	72.53
OTCYMIST [7]	74.85	67.69	71.09
Inkam	52.21	58.12	55.00

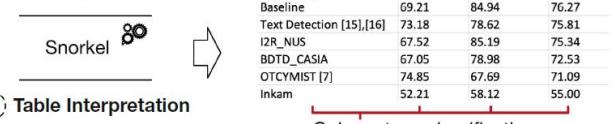
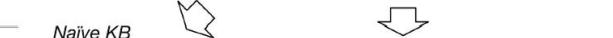
Input: PDF Figure



Assets

Method Name	Recall (%)	Precision (%)	F-score
USTB_TexStar	<b>82.38</b>	<b>93.83</b>	<b>87.74</b>
TH-TextLoc	75.85	86.82	80.96
I2R_NUS_FAR	71.42	84.17	77.27
<i>Baseline</i>	69.21	84.94	76.27
Text Detection [15], [16]	73.18	78.62	75.81
I2R_NUS	67.52	85.19	75.34
BDTD_CASIA	67.05	78.98	72.53
OTCYMIST [7]	74.85	67.69	71.09
Inkam	52.21	58.12	55.00

1 Table Extraction



2 Table Interpretation

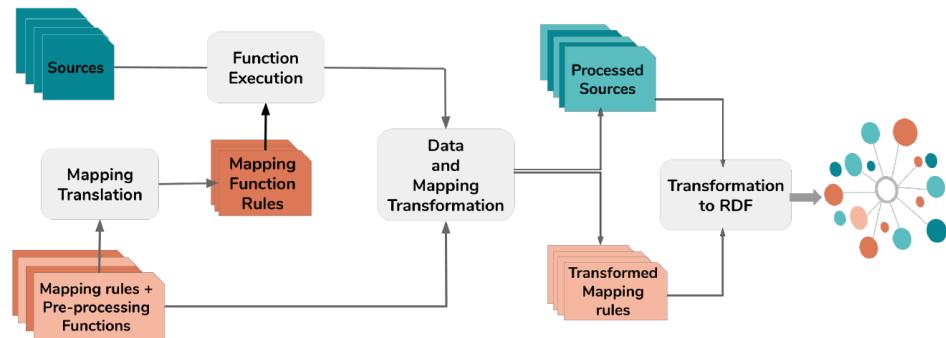


3 Entity Linking

Output: KB (with linked entities)

# FunMap: Efficient Execution of Functional Mappings for Scaled-Up Knowledge Graph Creation

- 概要
  - 関数ベースのマッピング言語の効率的なインタプリタ FunMapを提案
- 目的
  - データソース間の相互運用性を実現するために、効率的にマッピング言語を実行する
- 手法
  - 構文ベースの変換により、RML+FnOからFnOを具体的なRMLマッピングへ変換する
- 結果
  - 実行時間最大で1/18に短縮

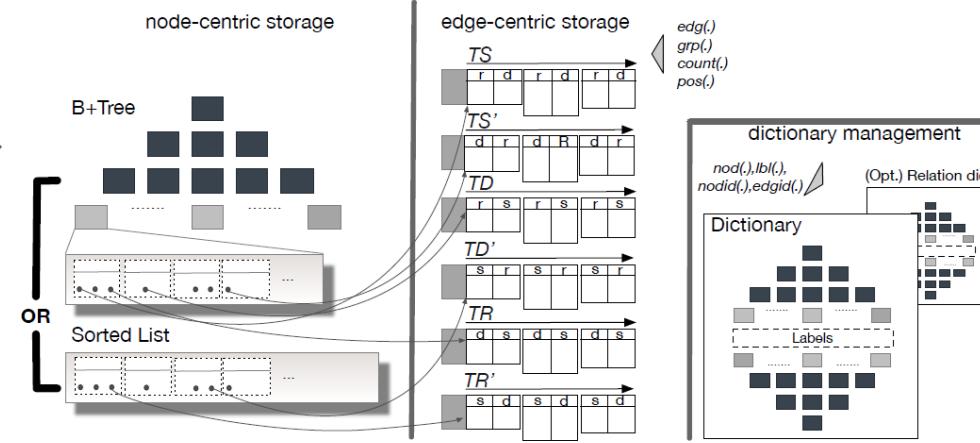


# **SESSION 7D: ROBUST AND SCALABLE MANAGEMENT OF SEMANTICS AND DATA**

# HDTCat: let's make HDT scale

- 概要
  - HDT(検索可能で省スペースなRDFのデータ形式)に対して2つのHDTファイルを結合するHDTCatを提案
- 目的
  - 巨大なHDTファイルの作成にはメモリが必要なので、個別に作成して結合することで境界条件を取り扱う
- 手法
  - 分割統治のアルゴリズムにより、省メモリに結合
- 結果
  - hdt2rdfよりもメモリ効率が良く、HDT-MRと同程度の効率でデータを結合することができた

# Adaptive Low-level Storage of Very Large Knowledge Graphs

- 概要
    - 大きなKG用の集中型ストレージアーキテクチャ Tridentを提案
  - 目的
    - KGはサイズが増え続けており、大きなKGに対して、複数タイプの計算を実行可能にしたい
  - 手法
    - B+Treeまたは配列によるノードへの高速アクセス
    - バイナリテーブルによるエッジの格納
  - 結果
    - $10^{11}$ 以上のエッジを持つKGを\$5Kのマシンで格納した
- 

# Cost- and Robustness-based Query Optimization for Triple Pattern Fragment Clients

- 概要
  - コストと堅牢性の両方を組み合わせたクエリプラン  
オプティマイザCROPを提案
- 目的
  - Triple Pattern Fragmentサーバーにおける効率的なクエリのため  
に、結合を考慮したクエリプランを提案
- 手法
  - ベストケースと平均ケースのコストの比較により、  
より堅牢なクエリプランを提案する
- 結果
  - 既存手法より全体実行時間とリクエストにおいて優れている

# SESSION 5D: KNOWLEDGE GRAPHS IN USE SESSION 6B: KNOWLEDGE DRIVEN NLP

砂川英一（株式会社東芝）

# 第54回SWO研究会

ISWC2020 サーベイ

5D KNOWLEDGE GRAPHS IN USE

6B KNOWLEDGE DRIVEN NLP

株式会社東芝 研究開発センター 砂川英一

2021.08.04

ISWC2020

---

## 5D KNOWLEDGE GRAPHS IN USE

# A Knowledge Graph for Assessing Aggressive Tax Planning Strategies (In-Use)

Niklas Lüdemann<sup>1</sup>, Ageda Shiba<sup>1</sup>, Nikolaos Thymianis<sup>1</sup>, Nicolas Heist<sup>1</sup>, Christopher Ludwig<sup>2,3</sup>, and Heiko Paulheim<sup>1</sup>

<sup>1</sup> Data and Web Science Group, University of Mannheim, Mannheim, Germany  
 {nluedema, agshiba, nthymian}@mail.uni-mannheim.de,  
 {nico, heiko}@informatik.uni-mannheim.de

## どんなもの？

多国籍企業の所属や子会社との関係をKG化。税対策の戦略を解析できるようにした。

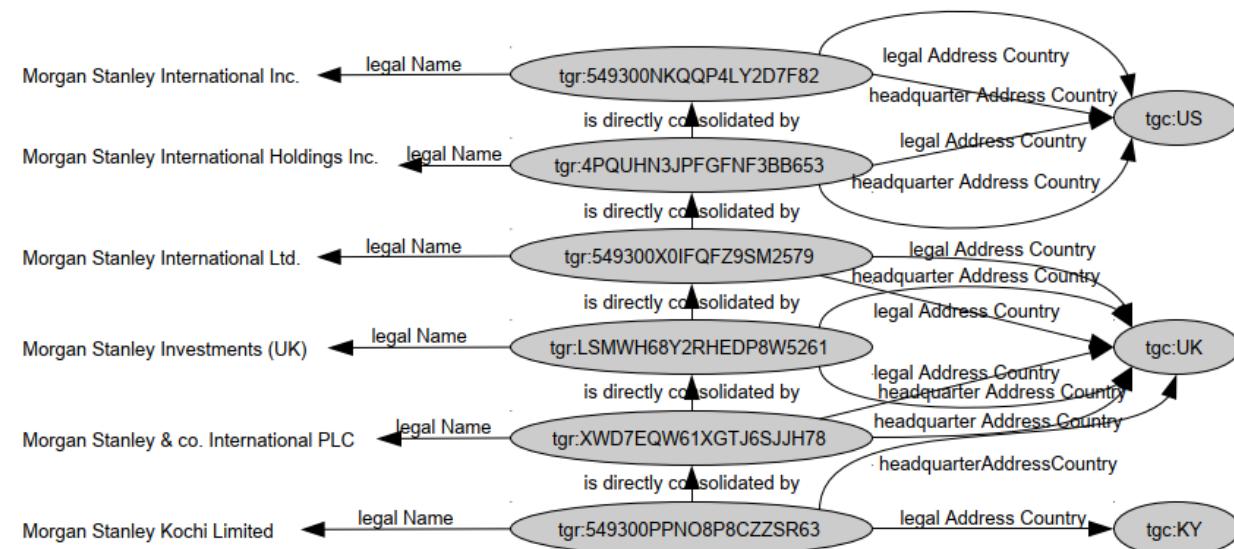
## 先行研究と比べて

- 多国籍企業の課税分析が可能。
- 企業間の親子関係の数が一桁多い。
- パブリック(クローズド版も開発予定)。

## 技術のキモはどこ？

- GLEIF(Global Legal Entity Identifier Foundation)のデータをKG化
- Double Irish with a Dutch Arrangement.など代表的な税対策、法人登録密度の高い都市などを抽出

## どんなKG/Framework？



## 参考資料

[A Knowledge Graph for Assessing Aggressive Tax Planning Strategies - YouTube](#)

# Turning Transport Data into EU Compliance while Enabling a Multimodal Transport Knowledge Graph (In-Use)

Mario Scrocca<sup>(✉)</sup>, Marco Comerio, Alessio Carenini, and Irene Celino

Cefri - Politecnico of Milano, Viale Sarca 226, 20126 Milan, Italy

{mario.scrocca,marco.comerio,alessio.carenini,irene.celino}@cefriell.com

## なんもの？

マルチモーダル輸送業のEU規制に準拠するためNational Access Pointsで輸送データを共有する必要がある。このデータ変換ソリューションを提案。

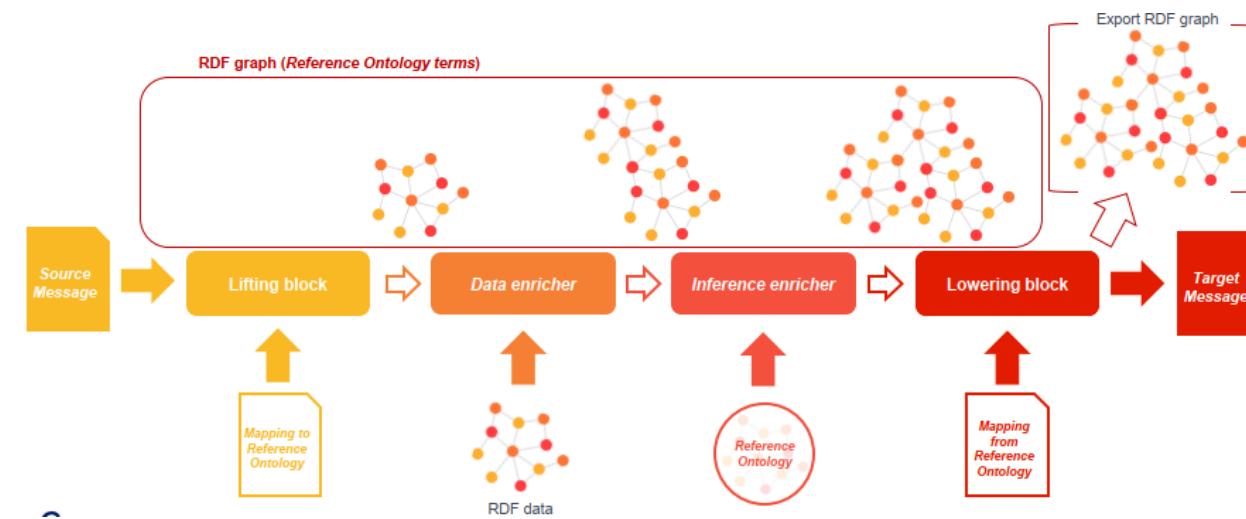
## 先行研究と比べて

既存のObject-relational mappingベースの変換手法よりメモリ消費が少なく、スケーラビリティが高い。

## 技術のキモはどこ？

- Chimeraフレームワーク：データ変換パイプラインを構築するためのモジュール群を用意
- データメッセージに対応するRDFグラフを添付して流し、各ブロックは共通のグローバルRDFグラフを参照しながら処理

## どんなKG/Framework？



## 参考資料

[Turning Transport Data to Comply with EU standards while Enabling a M… \(slideshare.net\)](#)

# Enhancing Public Procurement in the European Union through Constructing and Exploiting an Integrated Knowledge Graph (In-Use)

Ahmet Soylu<sup>1(✉)</sup>, Oscar Corcho<sup>2</sup>, Brian Elvesæter<sup>1</sup>, Carlos Badenes-Olmedo<sup>2</sup>, Francisco Yedro Martínez<sup>2</sup>, Matej Kovacic<sup>3</sup>, Matej Posinkovic<sup>3</sup>, Ian Makgill<sup>4</sup>, Chris Taggart<sup>5</sup>, Elena Simperl<sup>6</sup>, Till C. Lech<sup>1</sup>, and Dumitru Roman<sup>1</sup>

## どんなもの？

- EU活動における基盤となる“調達”KGと、これを公開・保管・統合・分析・可視化するためのAPIを開発
- 不正検知や文書検索ツールを開発

## 先行研究と比べて

- 国横断・言語横断で調達データと企業データをリンクしたプラットフォームやKGは無い。
- 多くは断片的であったり、記述要素に制限がある。

## 技術のキモはどこ？

- 主にOpenOpps(グローバル・オープンな入札データ)からデータを収集し、Open Contracting Data Standard(OCDS)ベースのオントロジーへマッピング。
- 統一化されたデータで入札不正の予測器を学習。

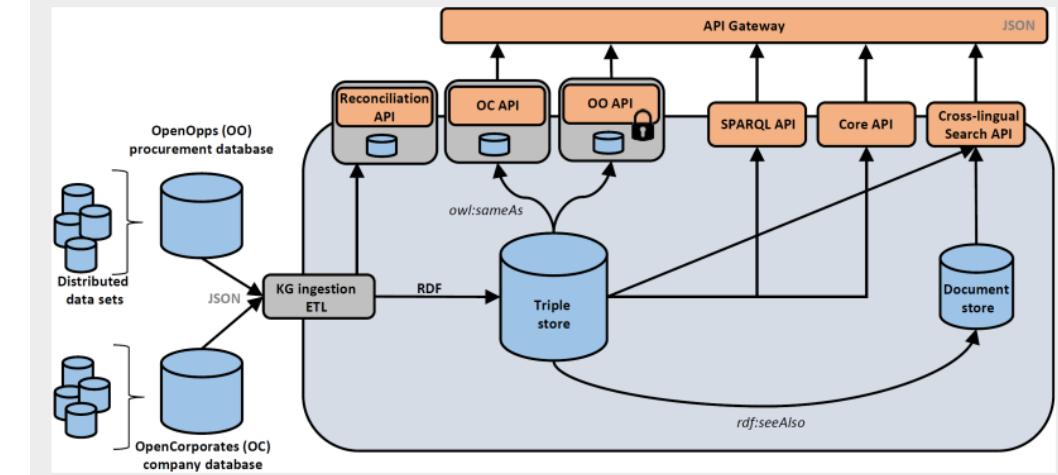
<sup>1</sup> SINTEF AS, Oslo, Norway

{ahmet.soylu,brian.elvesaeter,till.lech,dumitru.roman}@sintef.no

SINTEF is one of Europe's largest independent research organisations.

## どんなKG/Framework？

### TBFY platform: Architecture



## 参考資料

[TBFY - Enhancing Public Procurement in the EU through an Integrated Knowledge Graph - YouTube](#)

ISWC2020

---

## 6B KNOWLEDGE DRIVEN NLP

# Leveraging Semantic Parsing for Relation Linking over Knowledge Bases (Research)

Nandana Mihindukulasooriya<sup>(✉)</sup>, Gaetano Rossiello, Pavan Kapanipathi, Ibrahim Abdelaziz, Srinivas Ravishankar, Mo Yu, Alfio Gliozzo, Salim Roukos, and Alexander Gray

## どんなもの？

- 狙いはKBQA.
- SLING: Abstract Meaning Representation(AMR)とdistant supervisionを用いて意味解析→リンク生成するフレームワークを提案.

## 先行研究と比べて

- 自然言語の曖昧さを軽減できる
- 質問文とKB内のラベルとのギャップを埋める

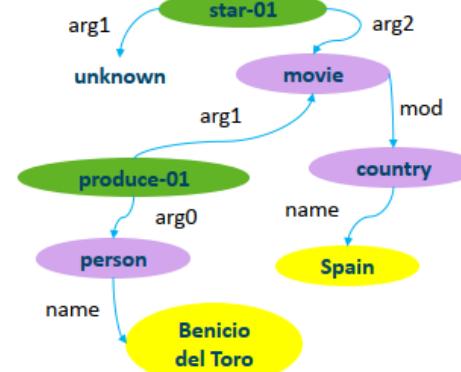
## 技術のキモはどこ？

- AMR-triple : 自然言語文とKG要素を取り持つ中間的なグラフを生成→対応するKG要素や知りたい解の型を予測
- Distant Supervision : Wikipediaの文とDBPediaのトリプルの対応関係を用いて、表現とKG要素の関係を学習

IBM Research, T.J. Watson Research Center, Yorktown Heights, NY, USA  
 {nandana.m,gaetano.rossiello,ibrahim.abdelaziz1,  
 sriini,alexander.gray}@ibm.com,  
 {kapanipa,yum,gliozzo,roukos}@us.ibm.com

## どんなKG/Framework？

Who is starring in Spanish movies produced by Benicio del Toro?



Triple 4

Triple 1  
 AMR predicate: produce-01.arg1.arg0  
 KB relations: [('dbo:producer', 3.60),  
 ('dbo:executiveProducer', 0.71),  
 ('dbo:manufacturer', 0.21)]  
 subject:  
 KB type: 'dbo:Film'  
 object:  
 KB URI: 'dbr:Benicio\_del\_Toro',  
 KB type: 'dbo:Person'

Output (Triples)

## 参考資料

[Leveraging Semantic Parsing for Relation Linking over Knowledge Bases \(slideshare.net\)](https://www.slideshare.net/ibmresearcher101/leveragingsemantic-parsing-for-relation-linking-over-knowledge-bases)

# A Novel Path-based Entity Relatedness Measure for Efficient Collective Entity Linking (Research)

国立情報学自動制御研究所

Cheikh Brahim El Vaigh<sup>1</sup>, François Goasdoué<sup>2(✉)</sup>,  
Guillaume Gravier<sup>3</sup>, and Pascale Sébillot<sup>4</sup>

## どんなもの？

- 文内のエンティティを、KB要素に結び付けたい。経路ベースでエンティティの関連性を評価する手法ASRMPmを提案。
- 3つの要件：Semanticsが明確、計算コストが合理的、推移律を扱えるを考慮

## どんなKG/Framework？

？？？  
(図か具体例が見たい…)

## 先行研究と比べて

従来手法（Weighted Semantic Relatedness Measure）は、Semanticsを重み変えて関係を評価。要件のうち推移律が満たせていなかった。

## 技術のキモはどこ？

対応するエンティティの候補は、同じ対象について述べていそうな他文からの候補も考慮して、絞り込む。

## 参考資料

？？？

# Leveraging Linguistic Linked Data for Cross-lingual Model Transfer in the Pharmaceutical Domain (In-Use)

Jorge Gracia<sup>1(✉)</sup>, Christian Fäth<sup>2</sup>, Matthias Hartung<sup>3</sup>, Max Ionov<sup>2</sup>,  
 Julia Bosque-Gil<sup>1</sup>, Susana Veríssimo<sup>3</sup>, Christian Chiarcos<sup>2</sup>,  
 and Matthias Orlikowski<sup>3</sup>

## どんなもの？

- 医薬分野での感情分析に向け、感情分類の学習モデルを言語をまたいで転移させる
- そのためのLLODとしてApertium RDFを利用

## 先行研究と比べて

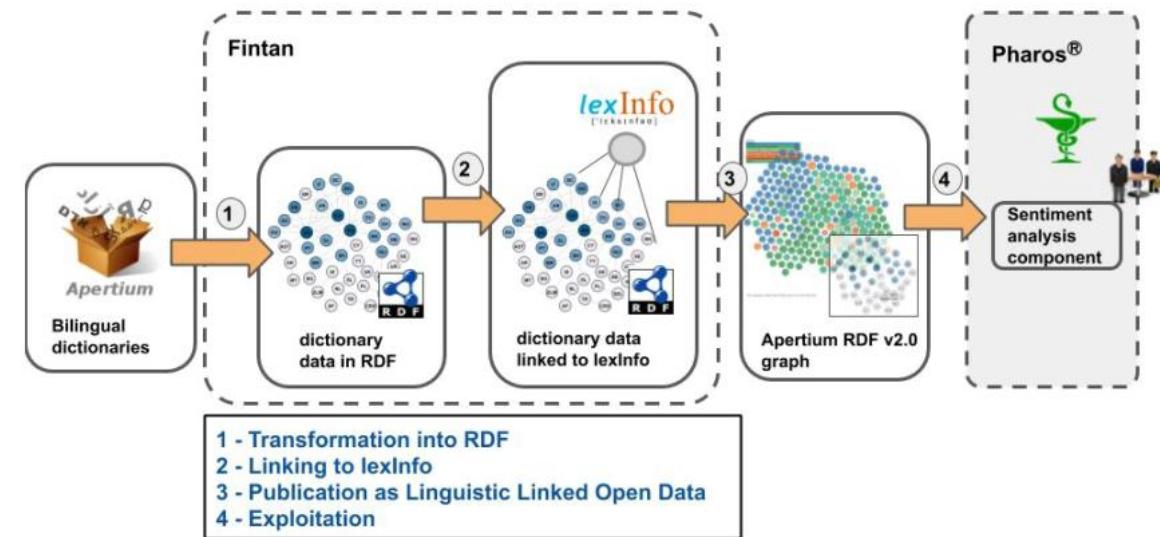
Apertium (a free/open-source machine translation platform) のデータをLD化  
 →他データとの連携が容易化、  
 元データの拡張がダイレクトにNLPアプリの改善に繋がる

## 技術のキモはどこ？

- BLSE (Bilingual Sentiment Embeddings)と名付けた転移学習を行い、ある言語で学習した感情の分類器を別言語で利用することで、再学習を不要にした
- 部分的にわかっている単語の対訳ペアを利用

<sup>1</sup> Aragon Institute of Engineering Research, University of Zaragoza, Zaragoza, Spain  
 {jogracia,jbosque}@unizar.es

## どんなKG/Framework？



## 参考資料

？？？

# SESSION 1B: KNOWLEDGE GRAPHS SESSION 4B: DATA MINING AND KNOWLEDGE-DRIVEN ML

勝島修平 (東京都市大学大学院 総合理工学研究科情報専攻)

# 第54回SWO研究会 ISWC2020サーベイ発表

- SESSION 1B: KNOWLEDGE GRAPHS
- SESSION 4B: DATA MINING AND KNOWLEDGE-DRIVEN ML

勝島修平

東京都市大学大学院総合理工学研究科情報専攻

# Refining Node Embeddings Via Semantic Proximity

(Research)

## 背景

random walkによるノード埋め込み手法

## 問題

意味的に豊富なネットワークの知識を十分に考慮していない

複雑な入力や利用不可能なノードタイプが必要

## 提案

述語間の意味的関係性を考慮したbiased random walkの提案

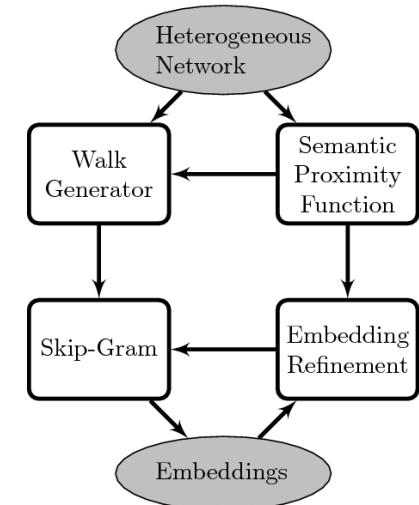
意味的な近接の概念を利用した埋め込みの改良

- ドメイン上の述語を利用
- penalty関数とsemantic近接性に基づく改良

## 結果

ノードクラスタリング：すべてのデータセットでベースラインを上回る

ドメインへの一致率：Biased-RDF2VecやRDF2Vecを上回る精度



# Enriching Knowledge Bases with Interesting Negative Statements (SC)

## 背景と問題

知識ベースは肯定的な情報しか保持しない

## 提案

KB上に否定的文脈を導入

否定的文脈の自動コンパイル手法を提案

➤ Peer-based statistical inferences:類似エンティティ(peer)の  
関係性から推論

➤ Pattern-based query log extraction:ログ抽出手法

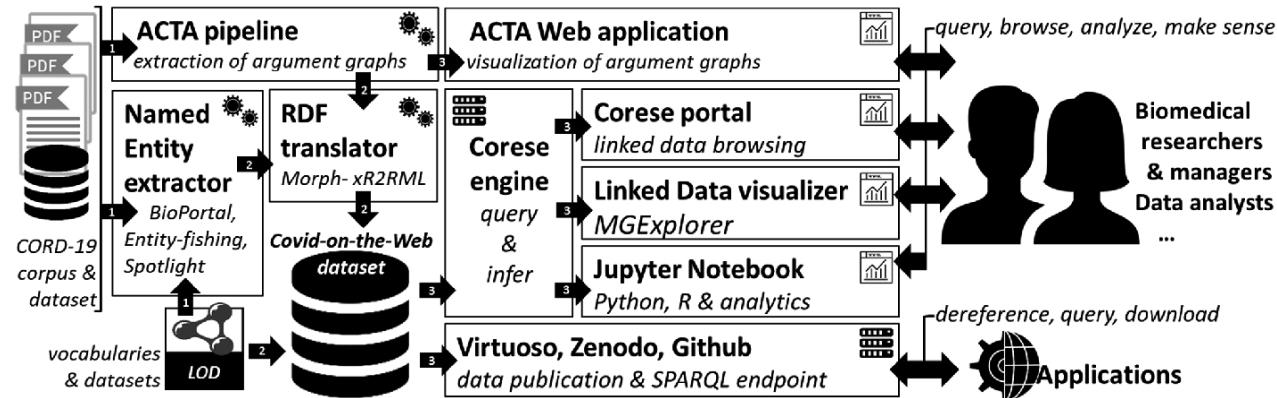
## 結果

両アプローチとも補完的な可能性を示唆

# Covid-on-the-Web: Knowledge Graph and Services to Advance Covid 19 Research(RS)

## 背景と問題

COVID-19関連の文献に対し、生物医学研究者が素早くアクセスし、照会し、理解できる枠組みが必要



## 提案

プロジェクトにおいて作成されたRDFデータセットとソフトウェアリソースの報告

- CORD-19コーパス (DBpedia, Wikidata, BioPortal等語彙にリンクされた2つのナレッジグラフ)
- ACTAによるCORD19ナレッジグラフの生成 (グラフの抽出と視覚化の自動ツール) Corese semantic web platform, MGExplorer, JupyterNotebook

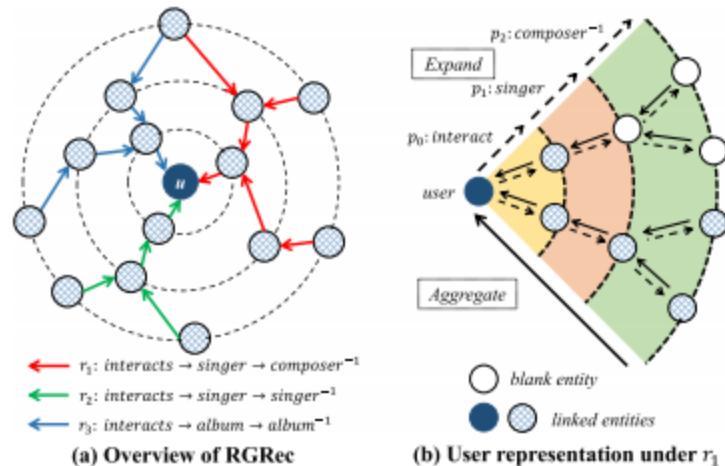
# Rule-Guided Graph Neural Networks for Recommender-Systems (Research)

## 背景

レコメンドシステムでの協調フィルタリングによって起こされるコールドスタート問題に対してナレッジグラフが採用されている

## 問題

ナレッジグラフ上の直接的な関係のみをモデル化



## 提案

ルール学習とGNNを組み合わせたRGRecによって、グラフ上の間接的に接続された関係をモデル化

- <1>ルール学習 高品質な推論規則発見のための規則フィルタリング
- <2>GNN ユーザー表現の獲得

## 結果

クリックスルー率において、比較手法より優位を確認

# BCRL: Long Text Friendly Knowledge Graph Representation Learning (SC)

## 背景と問題

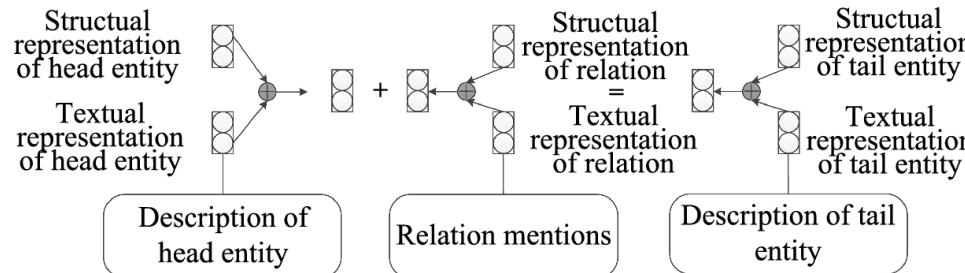
TransEをはじめとするTranslatingモデルは計算効率が高い一方で、トリプルの構造情報のみが考慮されるため、新しく追加されたエンティティまたは関係の少ないエンティティを正確に表すには不十分

## 提案

BERTとCNNを利用してテキストの様々な粒度を考慮

Relationに関連するトリプルすべてのエンティティ記述から取得される知識表現を強化するため、関係言及情報を導入

relation指向のattentionメカニズムを導入し、さまざまなコンテキストで、エンティティ記述の関連性の高い情報を考慮



## 結果

FreebaseとWordnetのベンチマークデータとテキストコーパスを利用して、リンク予想タスクでモデルを評価

提案モデルBCRL-AモデルがTransEより大幅に優れていることを確認

# Weakly Supervised Short Text Categorization using World Knowledge (Research)

## 背景と問題

- 短いテキストの分類タスクでは、テキストの希薄さと短さのために、従来の分類モデルを適用するとパフォーマンスが低下
- 教師あり学習ではラベルデータ作成の多大なコスト

## 提案

弱教師ありshort text classificationの提案(WESSTEC)

- 外部ナレッジグラフを利用した、特定のラベルなしトレーニングデータのラベル付けモジュール
- Wide&Deep Learningアプローチによる分類モデル

## 結果

- ラベルデータを必要としない
- 高精度で短いテキスト文書を分類  
(比較 SVM+tf-idf, BERT等)
- ラベルなし分類アプローチと比べても上回る性能  
(比較 LINE, Doc2Vec等)

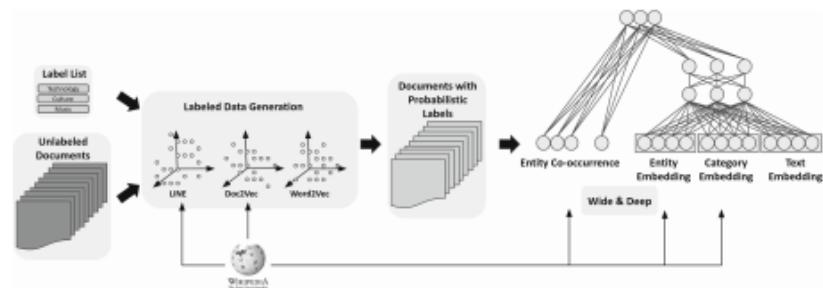


Fig. 1. The workflow of WESSTEC

# SESSION 9B: MACHINE LEARNING FOR THE SEMANTIC WEB

# SESSION 6C: ONTOLOGY ENGINEERING

鵜飼孝典（富士通）

# ISWC2020 紹介

2021年8月4日  
富士通株式会社  
鵜飼孝典

資料内の図はすべて原典からの引用です



shaping tomorrow with you

# SESSION 6C

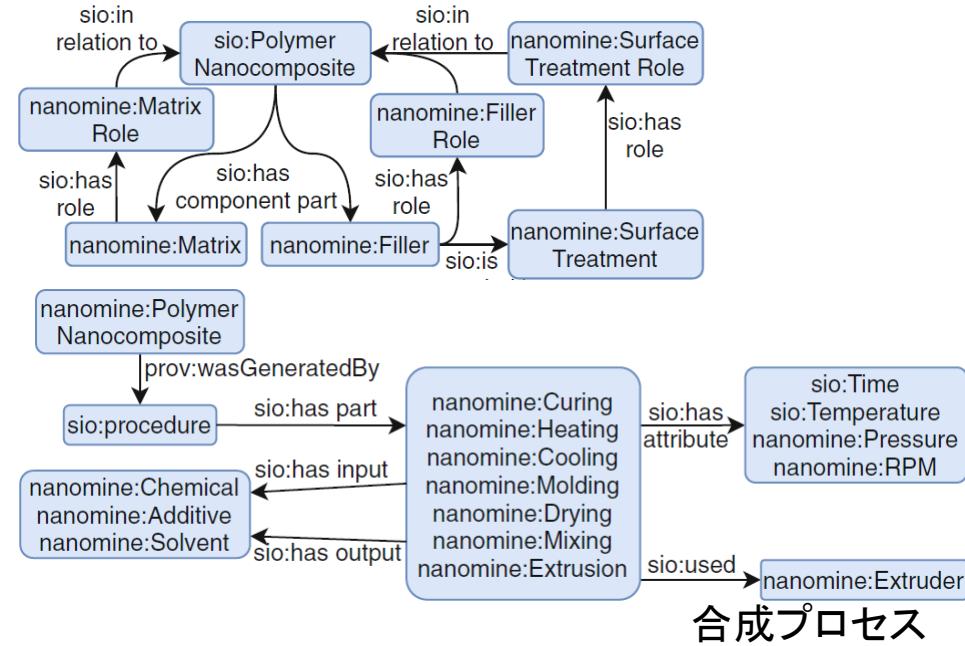
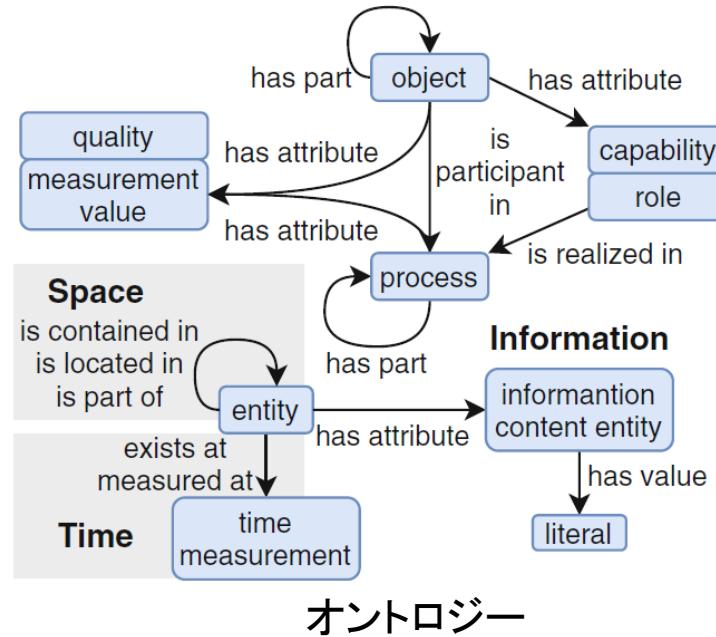
# ONTOLOGY ENGINEERING

# Nanomine: A Knowledge Graph for Nanocomposite Materials Science (RS)

FUJITSU

- 発表ビデオ : <https://www.youtube.com/watch?v=zy5lZxklja4>
- リソース : <http://nanomine.org>

高分子の合成



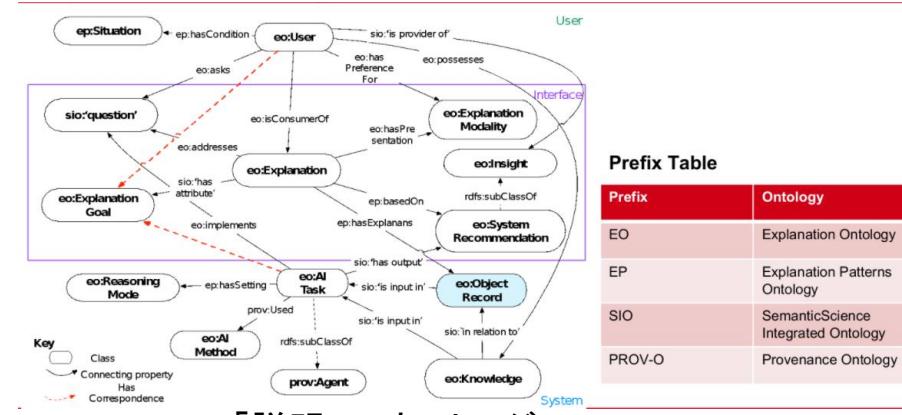
# Explanation Ontology: A Model of Explanations for User-Centered AI (RS)spotlight paper

FUJITSU

- Slide : <https://drive.google.com/file/d/1gCjdqguDVTi1z3iKDE-s60l0GYSFgAIA/view>
- Video : <https://tetherless-world.github.io/explanation-ontology/presentations/>
- リソース : <https://tetherless-world.github.io/explanation-ontology>

Explanation Type	Example
Case-based	What other situations with complex patients have had this recommendation applied?
Contextual	What broader information about the current situation prompted you to suggest this recommendation now?
Contrastive	Why administer this new drug over the one I would typically prescribe?
Counterfactual	What if the patient had a high risk for cardiovascular disease? Would you still recommend the same treatment plan?
Everyday	What are the signs I should be careful to check for in this case?
Scientific	What is the biological basis, particularly the evidence, for this recommendation?
Simulation based	What would happen if we prescribe this drug to the patient?
Statistical	What percentage of similar patients who received this treatment recovered?
Trace based	What steps were taken (rules were fired) by the system to generate this recommendation?

説明の種類と例

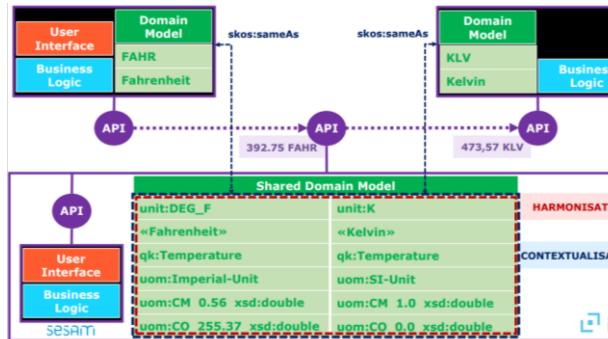


「説明」のオントロジー

# Revisiting Ontologies of Units of Measure for Harmonising Quantity Values – a Use Case (In-Use)

FUJITSU

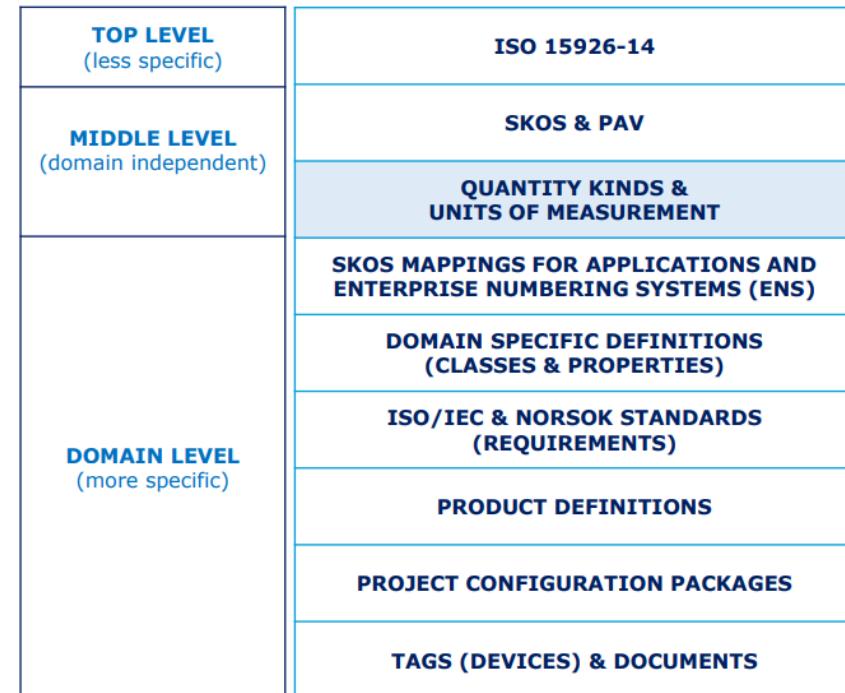
## ■ 複数の単位系を表現するオントロジーをix3モデルを用いて融合させる



単位変換

OMとQUDTの  
比較

REQUIREMENT	KEY WORDS	OM	QUDT
R01	Public License	+	+
R02	Actively maintained	+	+
R03	Coverage	+	+
R04	OWL 2 Direct Semantics	+	-
R05	(Web) Protégé & OWL API	+	-
R06	Efficient OWL reasoning (HermiT)	-	-
R07	Ontology design patterns	*	*
R08	Dimensional analysis & SPARQL	-	+
R09	Compatibility with ISO 15926-14	+	*
R10	Compatibility with SKOS	-	-
R11	Modular structure	-	+



ix3モデル



shaping tomorrow with you

# SESSION 9B

# MACHINE LEARNING FOR THE SEMANTIC WEB

# Relation-Aware Entity Alignment for Heterogeneous Knowledge Graphs (SC)

FUJITSU

■ 実装 : <https://github.com/StephanieWyt/RDGCN>

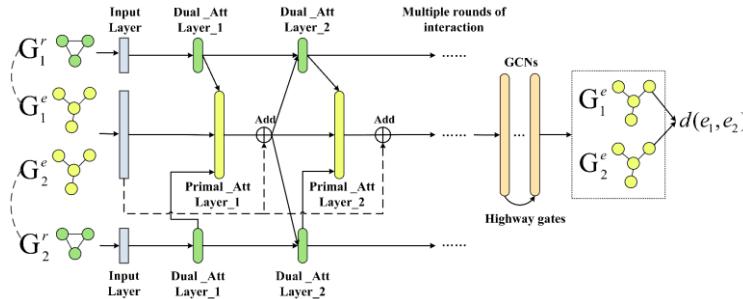
■ 想定タスク : entity alignment

■ 貢献

■ 既存技術 : Dual-Primal Graph CNN (DPGCNN) [Monti et al., 2018]



■ 概要説明図



図は原典から引用

# Multi-view Knowledge Graph Embedding for Entity Alignment (SC)

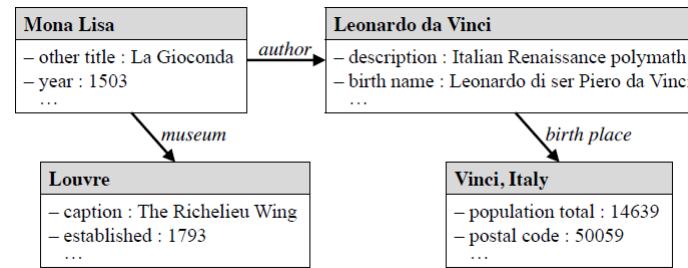
FUJITSU

■ 実装 : <https://github.com/nju-websoft/MultiKE>

■ 想定タスク : entity alignment

■ 貢献 :

- エンティティに対して名前、関係、属性のそれぞれの観点で埋め込みベクトルを定義
- 3つの観点の埋め込みベクトルの組み合わせ方を3通り定義
- BoostEAなどの既存の関係だけを定義している手法に対して、好成績

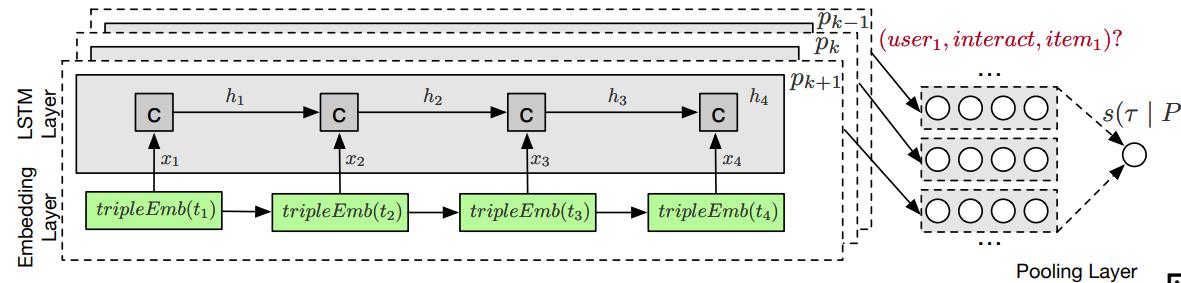


図は原典より引用

# Learning Triple Embeddings from Knowledge Graphs (SC)

FUJITSU

- 実装: <https://www.dropbox.com/s/en5ia4qk4o2b2us/T2Vec.zip?dl=0>
- 想定タスク: Triple Classification、Triple Clustering
- 貢献: 3つ組を単位として、埋め込みベクトルを定義する
  - 1つの3つ組を1つのノード、2つの3つ組でノードを共有しているとき、その2つの三つ組みにエッジがあるようなグラフに変換
  - 変換して得られたグラフについてランダムウォーク、SkipGramでベクトルを得る



図は原典から引用

# SESSION 2A: SHACL

隅田飛鳥 (AIP)

# ISWCサーべイ会 (Session2A & Session2B)

理化学研究所 革新知能統合研究センター 言語情報アクセスチーム

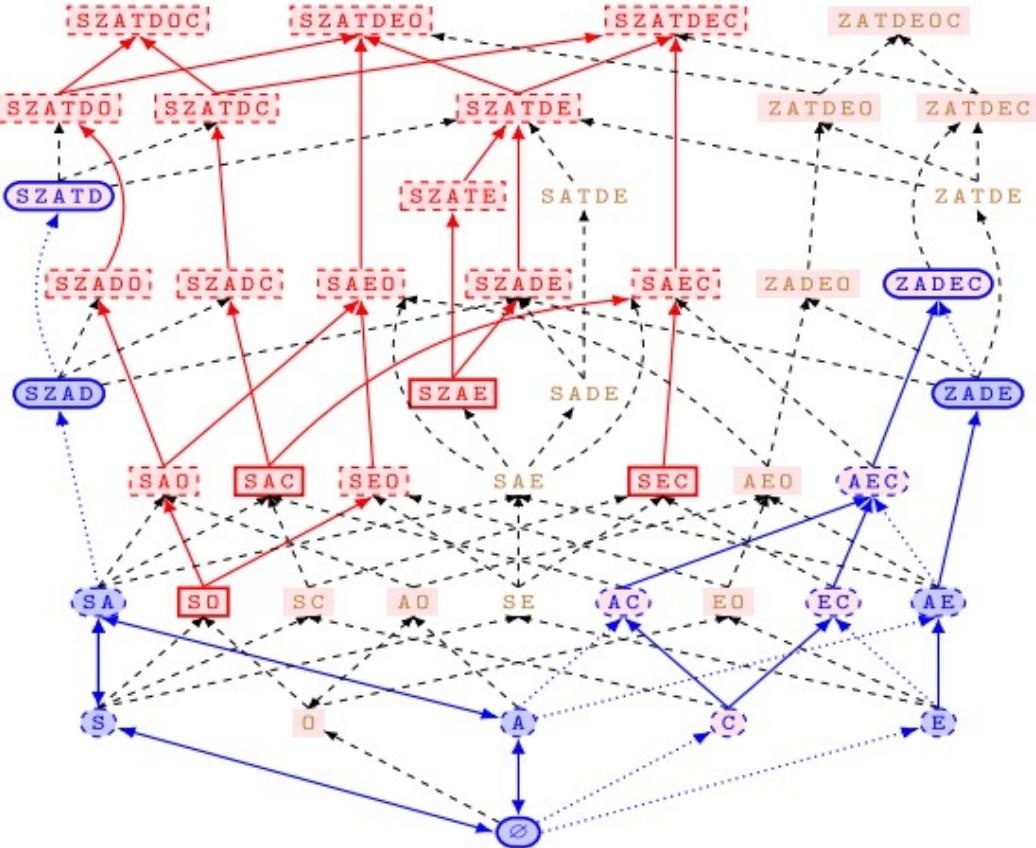
隅田飛鳥

# Session2A: SHACL

- SHACL Satisfiability and Containment (Research), Paolo Parietti, George Konstantinidis, Fabio Mogavero and Timothy Norman
- Deciding SHACL Shape Containment through Description Logics Reasoning (Research) **spotlight paper**, Martin Leinberger, Philipp Seifer, Tjitzé Rienstra, Ralf Lämmel and Steffen Staab
- Domain-specific customization of schema.org based on SHACL (In-Use) **spotlight paper**, Umutcan Şimşek, Kevin Angele, Elias Kärle, Oleksandra Panasiuk and Dieter Fensel

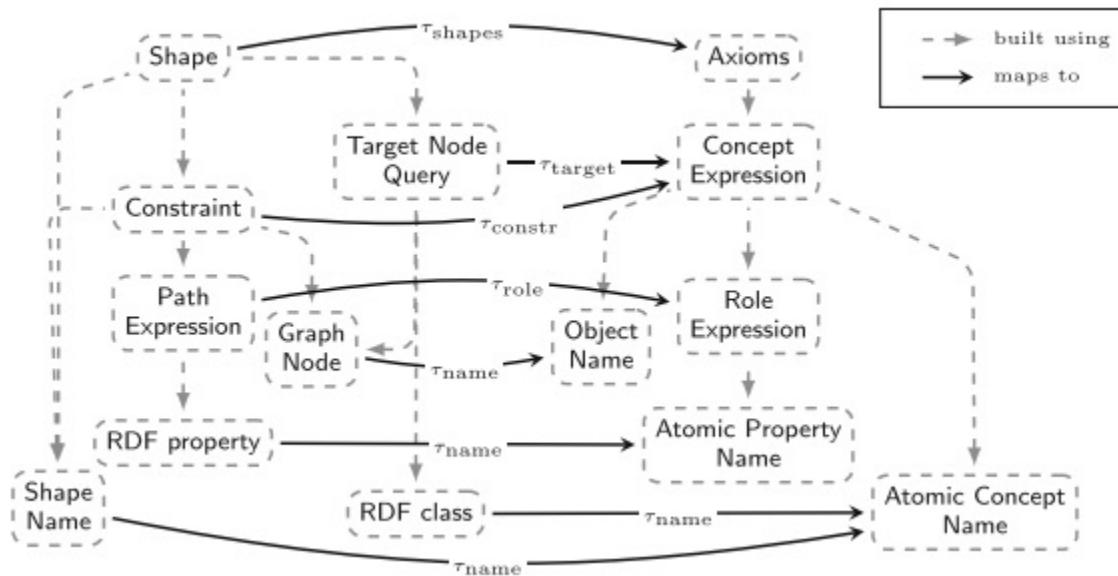
# SHACL Satisfiability and Containment

- 目的：SHACLのValidationの充足可能性問題と抑制を最適化して決定するSHACLの亜種言語SCLを作る
- 結果：SHACLをcounting quantifierとtransitive closureで拡張された一階述語論理に包括的な翻訳ができた



# Deciding SHACL Shape Containment through Description Logics Reasoning

- 課題: SHACLのShape同士の探索範囲の包含関係により、複数回、同じ場所を探索しなければいけない
- 目的: SHACLをDescription logicに変換して、Description logicを証明することで課題を避ける



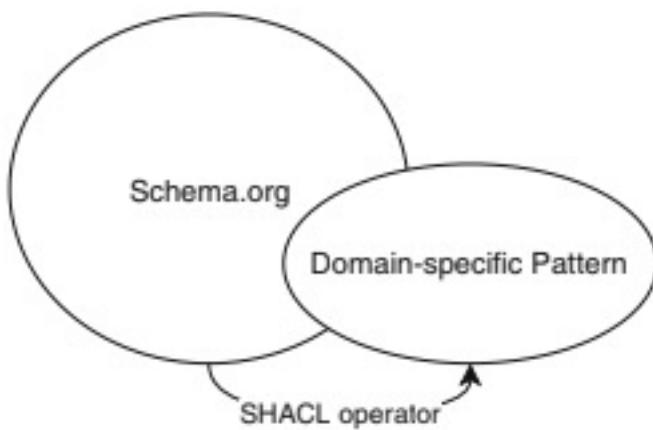
否定とパスの連続がなければ、完全に解けた。

SHACL Fragment	DL	Sound	Complete	Terminates
$\mathcal{L}$	$ALCOIQ(\circ)$	Yes	No	Not guaranteed
$\mathcal{L}^{\text{restr}}$	$SROIQ$	Yes	No	Yes
$\mathcal{L}^{\text{non-inv}}$	$ALCOQ$	Yes	Yes	Yes

# Domain-specific customization of schema.org based on SHACL

- 目的：schema.orgを特定のジャンルで、より詳細にしたい
- 手段：schema.orgと特定のジャンルで収集してきた知識をSHACLのValidationにかけることで、schema.orgに繋がるちょうどいい知識だけを付け加える

結果、プロがみて、75%、素人がみて20.7%がExcellentになった。素人の結果が低い理由は極端な評価が多かったため。



	$\bar{x}$	$\sigma$	$\tilde{x}$	Awful		Poor		Good		Excellent	
				$\bar{x}$	%	$\bar{x}$	%	$\bar{x}$	%	$\bar{x}$	%
DACH-KG	75	25	82.5	—	0	27.5	12.5	77.5	12.5	82.5	75
WSGT-students	49.82	16.5	50	20	3.4	36.5	17.2	48.23	58.6	70.41	20.7
ALL	55.27	21.24	55	20	2.7	35	16.2	49.86	48.6	76.45	32.4

# Session2B: KNOWLEDGE GRAPHS

- Crime event localization and deduplication (In-Use) Partially Reproduced, Federica Rollo and Laura Po
- Contextual Propagation of Properties for Knowledge Graphs: A Sentence Embedding Based Approach (Research), Pierre-Henri Paris, Fayçal Hamdi, Nobal Niraula and Samira Si-Said Cherfi
- Explainable Link Prediction for Emerging Entities in Knowledge Graphs (Research) Fully Reproducible, Rajarshi Bhowmik and Gerard de Melo

SESSION 3D: SEARCH, QUERY, INTEGRATION  
AND ANALYSIS

SESSION 2C: DATA MINING AND KNOWLEDGE-  
DRIVEN ML

小柳祐介（富士通株式会社）



shaping tomorrow with you

# ISWC2020サーベイ

富士通株式会社 小柳佑介

# SESSION 2C: DATA MINING AND KNOWLEDGE-DRIVEN ML

## ■ Diego Moussallem<sup>1</sup>, Dwaraknath Gnaneshwar<sup>2</sup>, Thiago Castro Ferreira<sup>3,4</sup> and Axel-Cyrille Ngonga Ngomo<sup>1</sup>

spotlight paper

video: [https://www.youtube.com/watch?v=-IMERb\\_1y1U](https://www.youtube.com/watch?v=-IMERb_1y1U)  
GitHub: <https://github.com/dice-group/NABU>

- <sup>1</sup> Data Science Group, University of Paderborn, Paderborn, Germany
- <sup>2</sup> DL Group, Manipal Institute of Technology, Manipal, India [dwarakasharma@gmail.com](mailto:dwarakasharma@gmail.com)
- <sup>3</sup> Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil
- <sup>4</sup> Tilburg Center for Cognition and Communication (TiCC), Tilburg University, Tilburg, The Netherlands [tcastrof@tilburguniversity.edu](mailto:tcastrof@tilburguniversity.edu)

### ■ タスク: Natural Language Generation - RDF-to-text

- 最近はneural modelが注目されているが、既存手法の対象は英語のみ

### ■ 提案手法: マルチリンガルなNeural model

- encoder-decoder architecture: encoder: GAT, decoder: Transformer

### ■ 結果

- WebNLGデータセットでのベンチマークで、EnglishでBLEU 66.21 (new SOTA), multilingual scenarioでBLEU 56.04

## ■ Janna Omelianenko, Albin Zehe, Lena Hettinger and Andreas Hotho

- Julius-Maximilians-University Würzburg, Am Hubland, 97074 Würzburg, Germany

video: <https://www.youtube.com/watch?v=WeLFyMwbEpA>  
GitHub: <https://github.com/JohannaOm/REWEIGHT>

## ■ タスク：言語モデルによる、Common Sense KGの向上

- KGを使って強い言語モデルを作る研究は多いが言語モデルを使って強いKGを作る研究は無い

## ■ 提案手法：

トリプルを自然言語文に変換し、言語モデルを使用してperplexityなスコア(REWEIGHT)を算出

- REWEIGHTスコアを付与したConceptNetを使うことで、既存のConceptNet Numberbatchよりも良いembeddingsを作成

※ISWC2020の併設ワークショップに、Common Sense KGのチュートリアル  
CSKGs (<https://usc-isi-i2.github.io/ISWC20/>) があった(この著者らは関係していない)

# Enhancing Online Knowledge Graph Population with Semantic Knowledge

(Research Track) 

- D`elia Fern`andez-Ca˜nellas<sup>1,2</sup>, Joan Marco Rimmek<sup>1</sup>, Joan Espadaler<sup>1</sup>, Blai Garolera<sup>1</sup>, Adri`a Barja<sup>1</sup>, Marc Codina<sup>1</sup>, Marc Sastre<sup>1</sup>, Xavier Giro-i-Nieto<sup>2</sup>, Juan Carlos Riveiro<sup>1</sup>, and Elisenda Bou-Balust<sup>1</sup>

■ <sup>1</sup> Vilynx, Inc., Barcelona, Spain

■ <sup>2</sup> Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

## ■ タスク: KG Population (KGの増強)

■ クラスタ化されたニュースイベントからのKG Population

## ■ 貢献

- 1. 関係抽出モデルにエンティティタイプの知識を手法の提案  
→TACREDベンチマークとTypeREデータセット上のF1スコアを向上
- 2. 関係抽出手法にさらにデータ検証手法を追加 →precisionを向上
- 上記を組み合わせることで、非構造webデータからのオンライン学習において、データの妥当性が向上
- ベンチマークのためにTypeREとAggregatedNewsREデータセットを構築

TypeRE: [https://figshare.com/articles/dataset/TypeRE\\_Dataset/12850154](https://figshare.com/articles/dataset/TypeRE_Dataset/12850154)

AggregatedNewsRE: [https://figshare.com/articles/dataset/AggregatedNewsRE\\_Dataset/12850682](https://figshare.com/articles/dataset/AggregatedNewsRE_Dataset/12850682)

# SESSION 3D: SEARCH, QUERY, INTEGRATION AND ANALYSIS

# OBA: An Ontology-Based Framework for Creating REST APIs for Knowledge Graphs

(Resources Track) 

spotlight paper / fully reproduced

## ■ Daniel Garijo and Maximiliano Osorio

- Information Sciences Institute, University of Southern California, Los Angeles, USA

## ■ KGにアクセスするためのREST APIを作るためのフレームワーク

- オントロジエンジニアとWeb開発者でギャップがあり、後者にはオントロジーなどの知識が無い
- それらの知識が無くてもKGの中身にアクセスできるREST APIを自動で作成

## ■ OWLオントロジーからOpen API Specification(OAS)を自動生成

## ■ OASを基に、JSONでやりとりするREST サーバとテストを生成

# Extending SPARQL with Similarity Joins

FUJITSU  
(Research Track)

## ■ Sebastián Ferrada, Benjamin Bustos and Aidan Hogan

- Department of Computer Science, Universidad de Chile, Millennium Institute for Foundational Research on Data, Santiago, Chile

slide and video: <http://sferrada.com/talk/iswc2020/>  
GitHub: <https://github.com/scferrada/jenasi>

## ■ SPARQLにおける多次元のSimilarity Join機能

- ## ■ 三つのSimilarity Joinアルゴリズムを、二つのデータセット上で比較
- DBSimJoin (Similarity JoinをサポートするPostgreSQL)と比較して高速

### iSPARQL

```
SELECT ?pub1 ?pub2 ?sim WHERE {
  ?pub1 rdfs:label ?title1 ;
    opus:book_title ?booktitle1 .
  ?pub2 swrc:label ?title2 ;
    swrc:book_title ?booktitle2 .
  IMPRECISE {
    ?sim1 isparql:jac (?title1 ?title2).
    FILTER (?sim1 >= 0.5) .
    ?sim2 isparql:lev (?booktitle1 ?booktitle2).
    FILTER (?sim2 >= 0.5) .
    ?sim isparql:score (?sim1 ?sim2 0.6 0.4).
    FILTER (?sim >= 0.5)
  } ORDER BY DESC(?sim)
```

### This work

```
SELECT ?p1 ?p2 ?d WHERE {
  {?p1 wdt:P31/wdt:P279* wd:Q6256 ; #countries
   wdt:P4010 ?GDP1 ;
   wdt:P1081 ?hdi1 ;
   wdt:P30 wd:Q46 } #in Europe
  SIMILARITY JOIN ON (?GDP1 ?hdi1) (?GDP2 ?hdi2)
  TOP 1 DISTANCE sim:manhattan AS ?d
  {?p2 wdt:P31/wdt:P279* wd:Q6256 ; #countries
   wdt:P4010 ?GDP2 ;
   wdt:P1081 ?hdi2 ;
   wdt:P361 wd:Q12585 } } #in LATAM
```

# Dynamic Faceted Search for Technical Support exploiting Induced Knowledge

(In-Use Track) 

■ Nandana Mihindukulasooriya, Ruchi Mahindru, Md Faisal Mahbub Chowdhury, Yu Deng, Nicolas Rodolfo Fauceglia, Gaetano Rossiello, Sarthak Dash, Alfio Gliozzo, and Shu Tao

■ IBM Research, T.J. Watson Research Center, Yorktown Heights, NY, USA

■ 動的ファセット検索を使った、ITサポートのためのインタラクティブなVirtual Assistant構築

■ 質問を与えると動的ファセットが提示され、ユーザが関心を絞り込むことを可能に  
■ テキストやSemanticなリソース(Wikidata)から自動的に知識を獲得

■ 結果

■ TechQA Benchmark / private datasetにおいて、Baseline(ElasticSearch)より向上

# SESSION 10A: OPEN DATA SESSION 1D: SCHOLARLY DATA

龜田堯宙（国立歴史民俗博物館）

**ISWC 2020 サーベイ企画@SIGSWO64TH**  
**SESSION 10A: OPEN DATA**  
**SESSION 1D: SCHOLARLY DATA**

国立歴史民俗博物館 龜田堯宙

# **SESSION 10A:** **OPEN DATA**

- **ODArchive – Creating an archive for structured data from Open Data Portals (RS)**

- *Thomas Weber, Johann Mitlöhner, Sebastian Neumaier and Axel Polleres*

- タイトル通り、オープンデータポータルからクロールしてアーカイブを作った論文。
- ここにある <https://archiver.ai.wu.ac.at/> GitHub にソースも <https://github.com/websi96/datasetarchiver>
- データの収集のスケール化：Kubernetes によるコンテナ管理、Kubernetes と連携するロードバランサとして NGINX Ingress Controller を活用、MongoDB のシャーディング（データを複数のサーバに分散させる機能）の活用など。
- 各種APIの提供：SPARQL API で DCAT（データカタログ）語彙で検索できる。表形式のデータについては CSV on the Web vocabulary 語彙も使って、表のカラム情報を提供している。他にも /get/file/{URL} でクロールされたファイルを提供するなど、個別の機能のAPIを備えている。クロールの依頼も Private API として実装されている。
- 分析：APIを使って、データ横断で同じ値をもつカラムを探したり、地名や組織名といったカラムのタイプを推定したりした。

#Resource URLs	798,091	Media Type	Count
#Versions	4,833,271	text/html	187,196
#Domains	6,001	text/csv	116,922
#Portals	137	application/json	102,559
Latest Versions Corpus Size	1.2 TB	application/zip	93,352
Total Corpus Size	5.5 TB	application/xml	76,862

- **Transparent Integration and Sharing of Life Cycle Sustainability Data with Provenance (In-Use) / spotlight paper / partially reproduced**

- *Emil Riis Hansen, Matteo Lissandrini, Agneta Ghose, Søren Løkke, Christian Thomsen and Katja Hose*
- ライフサイクル持続可能性評価（LCSA：Life Cycle Sustainability Assessment）
  - “Finkbeinerら（2010）により提唱された指標。持続可能性をトリプルボトムラインに基づき、社会、経済、環境の側面から、持続可能性スコアとして算出する。例えば社会性は、既存の人間開発係数（HDI）、ジニ係数（国民所得分配係数）やグローバルコンパクト等により評価し、経済性はコスト等により評価する。一方、環境は、LCAで評価し、資源、温暖化、酸性化、毒性などを影響領域としている。さらに影響領域間の重み付け係数についても検討をしていることが特徴である。” from 高橋和枝, 佐々木重邦「社会性ライフサイクルアセスメント指標の研究動向」
  - これを横断的に検索できると、例えば、中国の鉄鋼のライフサイクルでどのように使われてるか、どのような環境負荷があるか、とかがSPARQLできる。Named Graph 情報を使うことでその情報の来歴もチェックできる。
  - <http://odas.aau.dk> から SPARQL できる
  - 様々なスキーマのCSV等で書かれているデータを取ってきて BONSAI Ontology にマッピングして RDF 化
  - ライフサイクルを扱う BONSAI Ontology の各要素を来歴を扱うオントロジ PROV-O のプロパティでつなげたモデルを提唱。

- **GeoSPARQL+: Syntax, Semantics and System for Integrated Querying of Graph, Raster and Vector Data. (Research) / spotlight paper**

- *Timo Homburg, Steffen Staab and Daniel Janke*

- たとえば、地図 자체がベクターで表されていて、区域ごとにメッシュで区切られた洪水の予測があって、建物の情報がRDFグラフで表されていたら、それらを組み合わせてクエリしたい。特にラスタデータ関連が弱いから、GeoSPARQLを拡張する。
- ラスタは基本的に CoverageJSON Literal で表現。
- 領域の関係を演算として定義、FILTER文などにつかえる。洪水時に通れる道（浸水10cm未満）のクエリが以下。

```
{"type": "Coverage", "domain": { "type": "Domain", "domainType": "Grid", "axes": { "x": { "values": [-10,-5,0] }, "y": { "values": [40,50] } } }, "referencing": [ { "coordinates": ["y", "x"], "system": { "type": "GeographicCRS", "id": "http://www.opengis.net/def/crs/EPSG/0/4979" } } ], "observedProperty": { "ranges": { "FloodAT": { "type": "NdArray", "dataType": "float", "axisNames": ["y", "x"], "shape": [2, 2], "values": [0.5, 0.6, 0.4, 0.6] } } }}
```

Listing 1: Coverage JSON Literal example

```
1 SELECT ?road WHERE {  
2   ?road a ex:Road ; geo:hasGeometry ?roadseg . ?roadseg geo:asWKT ?roadseg_wkt .  
3   ?floodarea a ex:FloodRiskArea ; geo2:asCoverage ?floodarea_cov .  
4   ?floodarea_cov geo2:asCoverageJSON ?floodarea_covjson .  
5   BIND(geo2:rasterSmaller(?floodarea_covjson,10) AS ?relfloodarea)  
6   FILTER(geo2:intersects(?roadseg_wkt,?relfloodarea))}
```

Listing 2: Use Case 1: Flood Altitude

# **SESSION 1D:** **SCHOLARLY DATA**

- **The OpenCitations Data Model (In-Use) *spotlight paper***
- *Marilena Daquino, Silvio Peroni, David Shotton, Giovanni Colavizza, Behnam Ghavimi, Anne Lauscher, Philipp Mayr, Matteo Romanello and Philipp Zumstein*

- Open Citation という運動がある。論文の引用関係をオープンにしようという運動。基本的に DOI-to-DOI の関係を機械可読の任意のフォーマットで公開。本論文はそこにデータモデルを適用して、より詳細な情報を伴わせようとしている。
- <http://opencitations.net/model> にモデルの詳細情報アリ。DCAT とか PROV とかを組み合わせている。

ID	Description
P1	A classification of the type of citation (e.g. self-citation).
P2	The bibliographic metadata of the citing and cited bibliographic entities (e.g. type of published entity, identifiers, authors, contributors, publication date, publication venues, publication formats).
P3	The bibliographic reference, typically found within the reference list of the citing bibliographic entity, that references a cited bibliographic entity.
P4	The separate identifiers of all the in-text reference pointers included in the text of the citing entity, that denote bibliographic references within the reference list.
P5	The co-occurrence of in-text reference pointers within each in-text reference pointer lists (e.g. [3,5,12]).
P6	The identifiers of structural elements (e.g. XPath of sentences, paragraphs, captions) that specify where, in the full text, an in-text reference pointer appears.
P7	The function or purpose of the citation (e.g. to cite as background, extend, or agree with the cited entity) to which each in-text reference pointer relates.
P8	Provenance information of the citation extraction process (e.g. responsible agents, data sources, extraction dates).

Table 1. Representational requirements of the OpenCitations Data Model

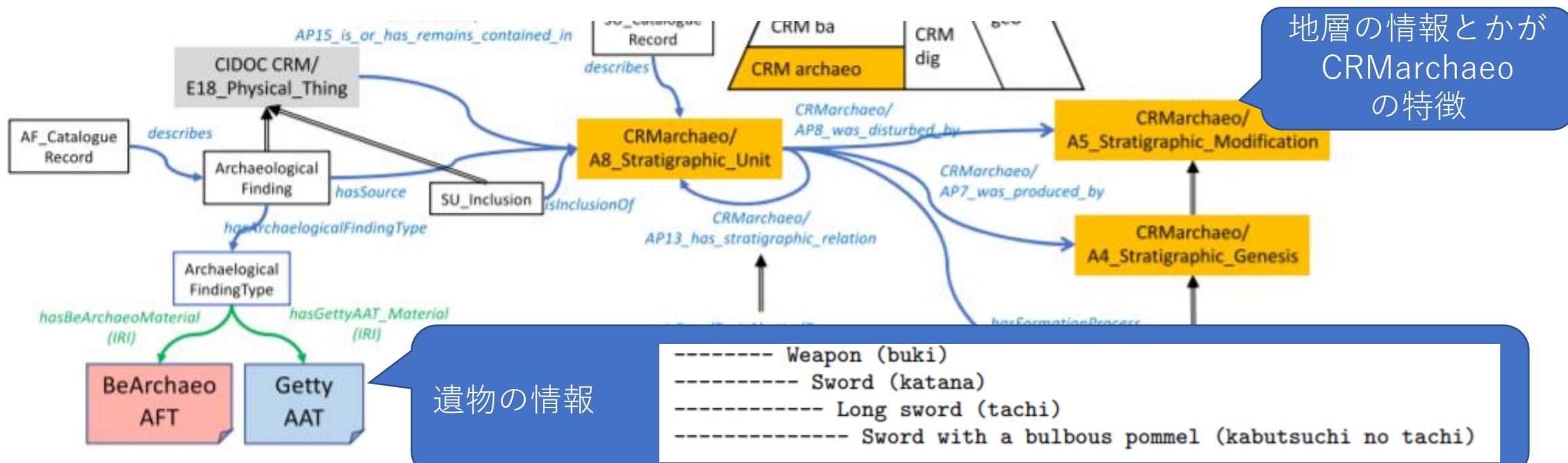
- P1: 引用タイプ (self-citationとか)  
 P2: 論文の書誌情報  
 P3: 論文の参考文献情報  
 P4: 参考文献のポインタ情報  
 P5: ポインタの共起情報  
 P6: 引用の生じている本文の位置情報  
 P7: 引用タイプ (cite as backgroundとか)  
 P8: この情報の来歴情報

他の語彙と比べてこれらすべてをカバーできることを確認している。

# • Linking ontological classes and archaeological forms (In-Use)

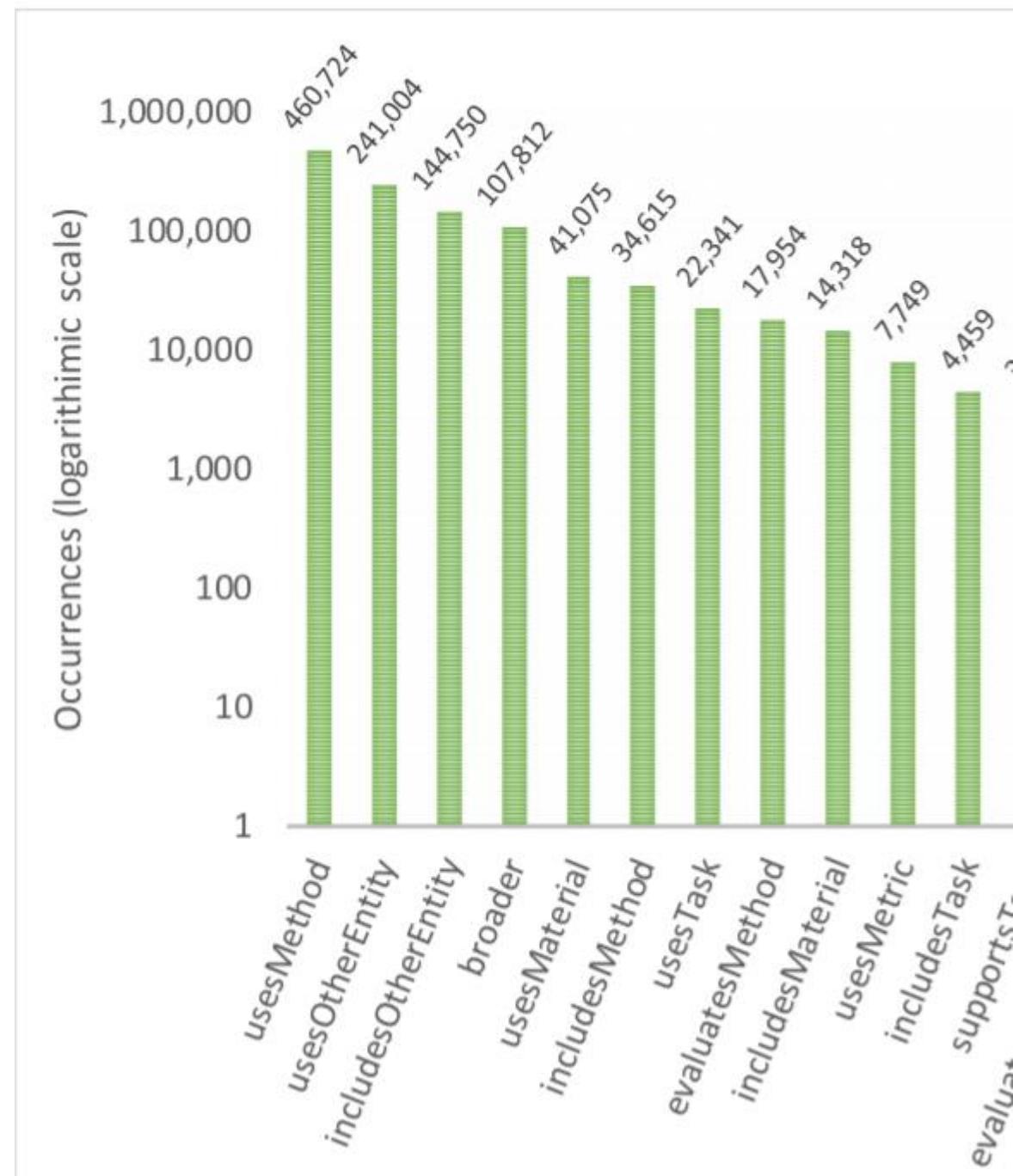
• Vincenzo Lombardo, Rossana Damiano, Tugce Karatas, and Claudio Mattutino

- 考古調査データをCRM Marchaeoで表現する仲立ちをする BeArcheo Ontology の紹介。
- 著者イタリア人だけど、考古データは具体的には岡山と島根の古墳だよ。
- CRM Marchaeo は CIDOC CRM の拡張。
- これをCMSに組み込んで Getty AATなどを値として選択しやすくしている。



- **AI-KG: an Automatically Generated Knowledge Graph of Artificial Intelligence (RS)**

- *Danilo Dessim, Francesco Osborne, Diego Reforgiato Recupero, Davide Buscaldi, Enrico Motta and Harald Sack*
- 論文群から知のネットワークを言語処理で抽出  
(...これはほぼ言語処理の論文です)
- <https://scholkg.kmi.open.ac.uk/> からアクセス
- {タスク、手法、指標、材料、その他}の5タイプとその間の27の関係
- 8種類の手法で一番複雑なのが一番性能がいい、基本DyGIE++が強い。
- 一番面白かったアイデア：信頼しきれん三重トリプルから信頼できるのを拾い上げるのに、信頼できるトリプルのConcat. Embeddingで学習して、主語と目的語から同じ述語を推定できれば信頼するという手法でF値約8%上げている。



# SESSION 7B: INDUSTRY 2

# SESSION 10B: QUESTION ANSWERING

浅野優（日立製作所）

# ISWC2020サーベイ

SESSION 7B: INDUSTRY

SESSION 10B: QUESTION ANSWERING

(株) 日立製作所 浅野 優

2021/8/4

# Unscripted Conversation driven through Knowledge Graph

## 【背景/課題】

- Accenture Technology Labsの取り組み
- チャットボットを構築するためには、人手でスクリプトを含む会話のモデリング（フローや教師データ）が必要であり、ルールやポリシーが変更されると、モデル変更が必要

## 【提案/実施内容】

- 人が理解しやすく、カスタマイズ、保守が容易な知識グラフを使用する方法を提案
  - ドメイン非依存のセマンティックスキーマを用いたドメイングラフの構築（2.1章）
  - ユーザ発話を処理し、結果を取得するためのセマンティック解析（2.2章）
- インタラクティブなダイアログシステムが構築でき、複雑なクエリも処理可能
- グラフに対してマルチホップクエリを実行可能

Keywords : Conversational AI、Knowledge Graph、Natural Language Processing

# SemNav: How Rich Semantic Knowledge Can Guide Robot Navigation in Indoor Spaces

## 【背景/課題】

- ・ インドのTATAグループにおけるロボットによるタスク実行にセマンティック知識を用いた取り組み
- ・ ロボットは、動的で不確実な環境（部分的に見えるオブジェクト、見えないオブジェクト、想定と異なった場所にあるオブジェクト等）においても、複雑なタスクを実行することが求められる

## 【提案/取り組み】

- ・ 豊富なセマンティックリレーションによる、ロボット、オブジェクト、および環境に関する知識を用いるために、ナビゲーションとオブジェクト検索問題固有の行動特性をリスト化し、SemNaviを作成
- ・ SemNaviは、主にSemNaviオントロジーとSemantic Navigationから成る。
  - ①SemNaviオントロジー：シードオントロジーと、画像シーンを処理して、意味的な関係として抽出したオブジェクトとキャプションから成る。抽出される関係には、Occlusion、Co-location、Location、Disjoint、Shape、Color等がある。
  - ②Sem Navigation：Semオントロジーに、解析シーンからのデータを入力後、人による検証をすることで、知識ベースを準備し、シーンの解析結果と組み合わせることで、ロボットの次の行動を導出する

Keywords : Ontology, Cognitive Robotics, Semantic Robot Navigation

# Using connected data to empower a financial services organization: Project Helix at UBS

## 【背景/課題】

- スイスの金融機関UBSにおいて、組織内でアプリケーション毎に管理されているデータをセマンティックを考慮して繋げることで、組織管理の合理化と強化をめざすProject Helixの紹介

## 【提案/取り組み】

- Cambridge Semantics社の製品Anzoを活用し、エンタープライズナレッジグラフを構築した
- Anzoは、データの取り込みから分析および配布までのend-to-endの機能を有していることから選択した。また、下記の2つの特徴がある。
  - 高速Query Lookupが可能なスケーラブルなin-memoryグラフストアであるAnzographを使用
  - 複数のSPARQLクエリを実行可能
- 既存のリレーションナルDBのメタデータ、スキーマなどを元に、OWLベースのドメインオントロジーを構築し、RDF形状に基づくマッピングモデルと方法論を開発した。
- 2つのケーススタディを紹介
  - 既存のテクノロジー資産を管理するDBを元にしたオントロジーの構築
  - RDF形状を用いたインスタンスデータのマッピング

# Knowledge Graphs and Creative Applications

## 【背景/課題】

- アイルランドのAccenture Labsの取り組み
- 食品や飲料などの消費財を提供する企業は、タイムリーに市場を満足させるイノベーションを起こすことが難しい（例. 健康的なスナックバーが市場動向で求められている時、健康上の利点があるチョコレートと組み合わせられる材料が必要となったりする）
- 食品ビジネスのプロフェッショナルの方が、新しい予期しない組合せである、斬新で興味深いフレーバーを見つけるための支援をし、製品が市場に出るまでの時間を短縮することが目的

## 【提案/取り組み】

- 新しい食品のアイデアを考えるプロセスで、知識グラフ、グラフエンベディングを活用するシステムを構築。
- 公開されている複数のデータから作成した、食材、レシピ、栄養素、化合物、フレーバーに関する知識グラフを使用。
- 食材の組み合わせに対して、「Surprise（驚き）」、「Pleasantness（快適さ）」、「Novelty（新規性）」を算出し、出力。

Keywords : Computational Creativity, Knowledge Graphs, Graph Embeddings

# Adventures in the Art of Enterprise Artificial Intelligence Transformation

## 【背景/課題】

- 米国政府が資金提供する研究開発センターを管理し、防衛、医療、サイバーセキュリティなどの分野で政府機関をサポートするMITREの取り組み
- 再構築することなく、スケーラブルで持続可能で費用対効果の高い、人工知能を実装する方法が求められている

## 【提案/取り組み】

- 3年間のエンタープライズ向けAIでの取り組みから得た経験と教訓の成果として、MITER Embedded Intelligence Frameworkを紹介している
- オープンソースとセマンティックWeb技術を組み合わせて、構造化されていないテキストから洞察を抽出し、AIライフサイクルを管理し、AIアーティファクトを共有し、AIサービスを届ける
- アプリケーション例には、検索、出版物のレコメンド、チャットボット、ロボットプロセス音声対応の仮想アシスタント等がある

Keywords : Knowledge Graphs, SHACL, NLP, Machine Learning, RDF, Enterprise AI, SKOS, RDFS

# CASQAD – A New Dataset For Context-aware Spatial Question Answering

## 【背景/課題】

- ・ ドイツのVolkswagenとFraunhoferの研究所における、状況に対応するコンテキストアウェアな空間的な質問応答のための新しいデータセットに関する取り組み
- ・ フاكتoidの質問応答のタスクは、スマートフォンなどを用いてコンテキスト情報が急速に変化する場合に対応することが難しい（例. 目の前の石碑の建築家は誰？と質問される場合）
- ・ 従来のQ&Aシステムでは、時間的および空間的な質問には対応できない

## 【提案/取り組み】

- ・ 状況に対応する適切なクエリを作成するために、ユーザーの位置と移動方向に関する情報を表す、視覚的および空間的な参照を含む注釈付きの質問をコンテキストアウェア空間QAデータセット（CASQAD）を作成し、公開した
- ・ これらの質問は、Amazon Mechanical Turkを利用し、ドイツのハノーバー市の25のパノラマについて、400人から、5,232件を超える質問で収集され、品質を確保するための適切な手段を使用して、半自動で注釈を付けた
- ・ Googleストリートビューからのメタコンテキスト情報に応じて質問を拡充した
- ・ Volkswagenのデジタルアシスタントの研究分野で使用している。

Keywords : Datasets, Benchmark, Question Answering, Knowledge Graphs

# RuBQ: A Russian Dataset for Question Answering over Wikidata

## 【背景/課題】

- ロシアのITMO大学とUral Federal大学とJet Brainsという研究所における、質問応答のためのロシア語のデータセットに関する取り組み

## 【提案/取り組み】

- Wikidataを介した質問応答用の最初のロシアのデータセットであるRuBQを作成
  - 1,500件の質問、英語への機械翻訳、およびWikidataへの注釈付きのSPARQLクエリで構成される
  - データセットには、ロシア語と英語のラベルが付いた810万のエンティティを含む2億1200万のトリプルのWikidataのサンプルと、評価スクリプトが付属している
  - オンラインクイズの質問と回答のペアからなる大規模なコレクションに基に、自動フィルタリング、クラウドソーシングによるエンティティリンク、SPARQLクエリの自動生成、社内検証を組み合わせることで、効率的に作成された
  - セマンティックWeb、質問応答、およびセマンティック解析の分野の研究者の幅広いコミュニティで活用できる
- 300件のRuBQの質問には答えられないため、今後の課題になっている

Keywords : Knowledge base question answering, Semantic parsing, Evaluation, Russian language resources

# PNEL: Pointer Network based End-To-End Entity Linking over Knowledge Graphs

## 【背景/課題】

- ・ドイツのHamburg大学、Fraunhofer、Bonn大学における、エンティティリンクの精度向上に向けた取り組み
- ・質問応答システムでは、エンティティリンクが最初のステップとなることが多く、その中で、最初にスパンを検出し、次にエンティティの曖昧性の解消を行う場合があり、そのステップでのエラーが、全体的な精度を低下させてしまうことがある  
例。「Who founded Tesla?」の「Tesla」は、人物なのか、会社なのか、磁束密度の単位なのか

## 【提案/取り組み】

- ・知識グラフを介したポインターネットワークモデルに基づくend-to-endのエンティティリンクへの新しいアプローチPNEL (Pointer Network based Entity Linker) を提案
- ・Wikidataの知識グラフの3つのデータセット (WebQAP、SimpleQuestions、LC-QuAD) に対して評価を実施し、既存手法と比較したところ、2つのデータセットで精度の向上し、1つのデータセットで適合率は向上したが、再現率は低下した
- ・今後は、Transformerモデル等の他の手法を適用することを検討している

Keywords : Entity Linking, Question Answering, Knowledge Graphs, Wikidata

# SESSION 6D: INDUSTRY 1 SESSION 10D: IN-USE APPLICATIONS

長野伸一（東芝）

# **人工知能学会 第54回SWO研究会 ISWC2020サーベイ**

2021/08/04

株式会社東芝  
長野 伸一

# **Session 6D: INDUSTRY 1**

# A Context-aware Recommendation System for Mobile Devices

Samsung Research, University of Oxford

## ■ 概要

- ユーザのコンテキストを理解した推薦システムを開発

## ■ 課題

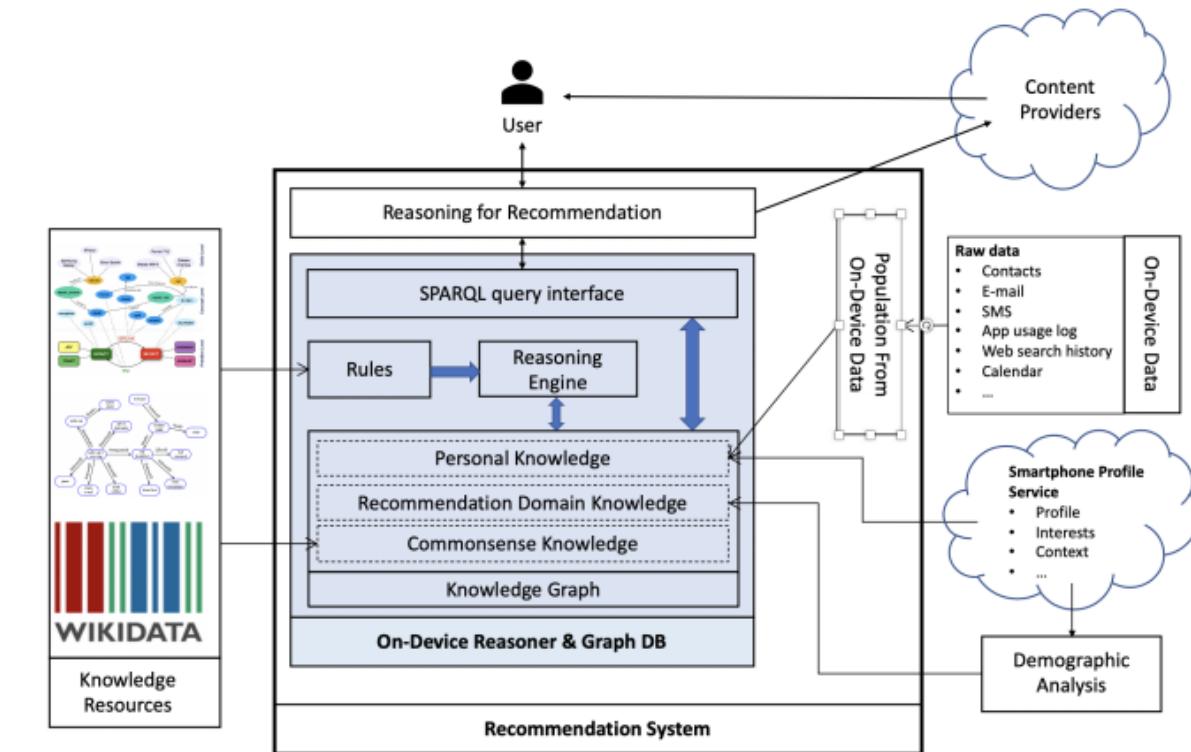
- ユーザのコンテキストの把握
- プライバシーへの懸念
- コールドスタート問題

## ■ 解決手段

- モバイル端末に保存されている豊富なデータを分析
- モバイル端末上で推論処理を実行
- 知識ベースを利用した推論型アプローチを採用
  - 実装にはRDFox組込版を採用。高速なクエリ応答、インクリメンタルな実体化をサポート

## ■ 評価・実績

- 推薦の処理性能はLUBMで評価、推薦の質はサムソン従業員による評価で有用性を確認
- 音声アシスタント搭載のサムソン製スマートフォンに採用予定



**Fig. 1.** On-device reasoning-based recommendation system.

# Semantic-based Quality of Service Management for Real-time WebRTC Streaming Service

Samsung Research Headquarter, Samsung Research in Bangalore

## ■ 概要

- スマートフォンのゲーム画面をLAN内のスマートTVにストリーミングするサービスにおけるQoS管理機構を開発

## ■ 課題

- WebRTCベースのストリーミングでは、スマートTVとスマートフォンが同じLAN内にあることを前提としているが、ユーザ体験の質を満たすには、あらゆる種類の遅延を考慮する必要がある

## ■ 解決手段

- WebRTCゲームストリーミングサービス上に開発したセマンティック層で、遅延に関する情報（ハードウェア情報、ネットワーク情報など）をRDFで管理し、ユーザー体験を最大化するコンフィグレーションを設定する

## ■ 評価・実績

- 製品搭載について記載なし

# An Unsupervised Framework for Semantics Driven Causal Explanations for Anomalies

IBM T.J. Watson Research Center

## ■ 概要

- 多変量の時系列センサーデータを利用した設備異常検知タスクにおいて、因果関係の局所的な説明を得る、教師なし学習手法（因果関係ベイジアンネットワーク技術）を開発

## ■ 課題

- ベイジアンネットワークにもとづくグラフ学習手法の多くはノード数増加に伴う計算量の問題がある

## ■ 解決手段

- Greedy hill-climbingによる構造学習、及び相互情報量に基づく評価指標を利用し、探索対象を有向リンク候補に限定することで計算量を削減
- グラフの構造変化が最大となる時点を異常発生時と特定し、この異常発生前後のグラフで異常を説明する  
※本論文中で“Semantics”が何を指すのか不明

## ■ 評価・実績

- 蒸気タービンの機械振動異常への適用評価で有用性を確認
  - 故障の発生時点を特定
  - 故障前後で変化が最大のセンサを特定

## ■ 概要

- ・質問応答システムAnuQAを構築

## ■ 課題, 及び解決手段

- ・データ補強：リンク予測等の下流タスクが性別,民族,宗教等の属性に依存しないよう,学習データの多様性を向上
- ・エンティティ再解決：知識グラフへの更新に対し, Dirichlet Hawkes過程を用いてテキスト類似性と時間的近さをモデル化
- ・オントロジーアライメント：Unified Hierarchical Label Setモデルを用いて異なるデータソースからの情報統合を自動化
- ・リンク予測の説明性：予測される新しいリンクを説明するために, 記述的説明を表すテキストを抽出
- ・意図理解：自然言語クエリと論理的推論を融合し, 分析的な問い合わせにおける暗黙の意図を解釈

## ■ 評価・実績

- ・顧客エンゲージメントや製品情報提供, 金融サービスなどに導入
- ・COVID19 India Question Answering Systemで試用可能  
<http://covid19-india-qa.mybluemix.net>

# Semantic ML for Manufacturing Monitoring at Bosch

Bosch CR, Karlsruhe Institute of Technology, Bosch Center for AI, University of Oslo

## ■ 概要

- ・ 製造プロセスの品質監視向けに、オントロジーを活用した機械学習（ML）のパイプラインSemMLを構築

## ■ 課題

- ・ 製造プロセスにおける機械学習活用の課題

1. 透明性：製造分野の専門家と機械学習の専門家による、ML品質モデルやMLの結果の合意に時間を要する
2. データ整備：データの統合には、多面的なドメイン知識、及びデータの複雑さへの理解が必要
3. ML品質モデルの汎用性：特定のデータセットとプロセス向けに構築されたモデルの転用には労力を要する

## ■ 解決手段

- ・ オントロジーを活用したMLパイプラインを構築
- ・ テンプレートによる製造知識、及び機械学習知識のデータ化
- ・ データへの意味情報のアノテーションにより、MLモデリングを自動化でき、モデルの説明性を向上

## ■ 評価・実績

- ・ ボッシュにおける溶接品質監視への適用評価で有用性を確認  
※知識データの構築と活用の“型”は整備されたが、手作業の要素がまだ多い印象

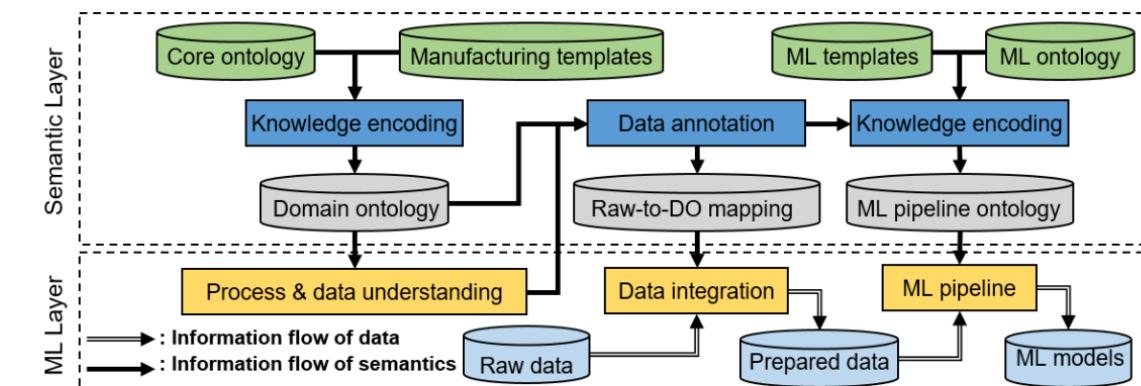


Fig. 1: Overview architecture of our approach and the SemML system.

# **Session 10D: IN-USE APPLICATIONS**

# Semantic Framework for Enabling Radio Spectrum Policy Management and Evaluation

Rensselaer Polytechnic Institute, The Rensselaer Institute for Data Exploration and Applications, Memory Based Research LLC, CACI International Inc., Air Force Research Laboratory

ダイナミック周波数共用(Dynamic Spectrum Access, DSA)技術  
従来、周波数帯ごとに管理されていた電波をデータベースで一元管理し、遊休の周波数帯域を事業者に割り当てるこを可能にする技術。

## ■概要

- DSAポリシーフレームワークに基づいてポリシーの管理や評価を支援するWhyisシステムを開発

## ■課題

- 電波は限られた資源であるため、その利用と管理は政府機関が規制・管理している。商用利用が可能な電波が増えてきており、電波管理の効率化が求められている

## ■解決手段

- DSAポリシーをOWLとPROV-Oによる機械可読な形式で記述・管理する。電波のアクセス要求に対して推論により評価を行い、アクセス可否の評価結果を提示する

## ■評価・実績

- シナリオに基づき作成したアクセス要求に対し、専門家の評価と推論による比較評価を実施

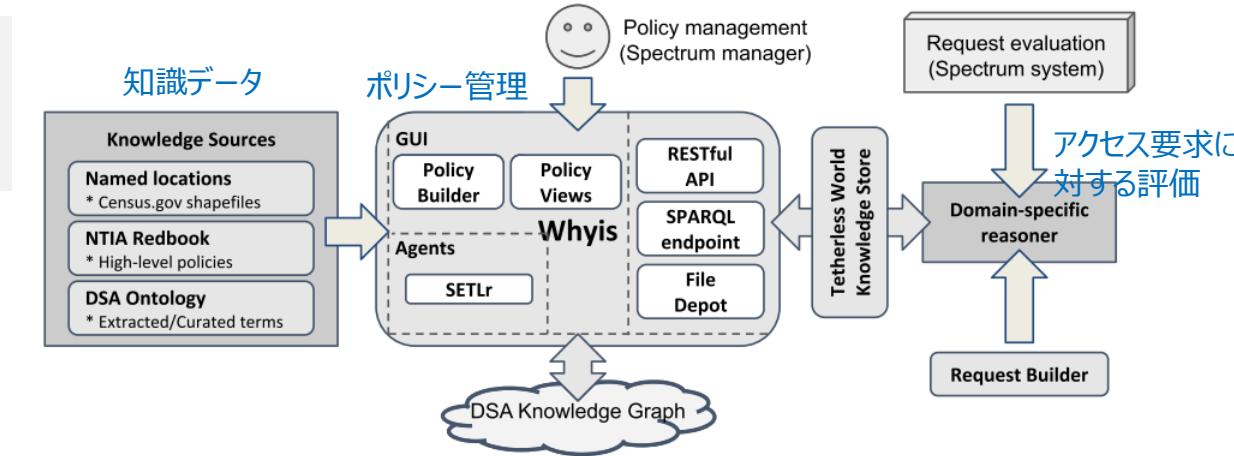


Fig. 1. DSA Policy Framework architecture  
フレームワークの概要

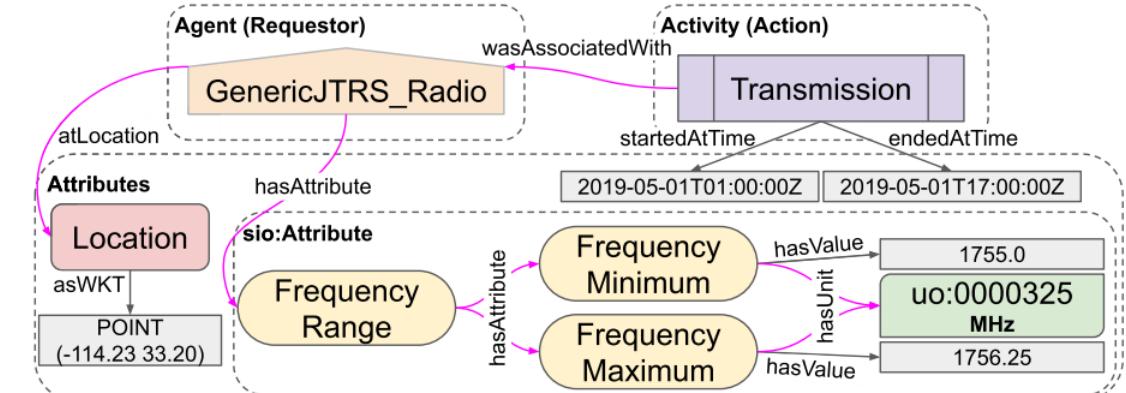


Fig. 3. The DSA request model and sample request  
DSAポリシーに対するアクセス要求記述の例

# AWARE: A Situational Awareness Framework for Facilitating Adaptive Behavior of Autonomous Vehicles in Manufacturing

BMW Group, Technical University of Munich, Karlsruhe Institute of Technology, Saint Joseph University of Beirut

## ■ 概要

- 自動車製造工場内AGV向け状況認識フレームワークを開発

## ■ 課題

- センサ情報だけではタイムリーかつ秩序ある動作を保証することは難しく、状況認識を補完する必要がある

自動車工場内でのケーススタディから得られた課題：1)AGVの俊敏性が有人車に比べて低い、2)交差点や狭い通路などの状況で衝突の可能性がある

## ■ 解決手段

- AGVの状況認識を強化する知識ベースのフレームワークを構築（オントロジー、ルール、推論器、ストリーム処理で構成）
- AGVの行動規則を作成（安全性の標準規格が十分に整備されておらず、工場のオペレーションをもとに作成）

## ■ 評価・実績

- 定量評価：約50万のRDFトリプルに対する推論の処理時間でスケーラビリティを確認
- 定性評価：Unityのシミュレーション環境でAGV行動を分析

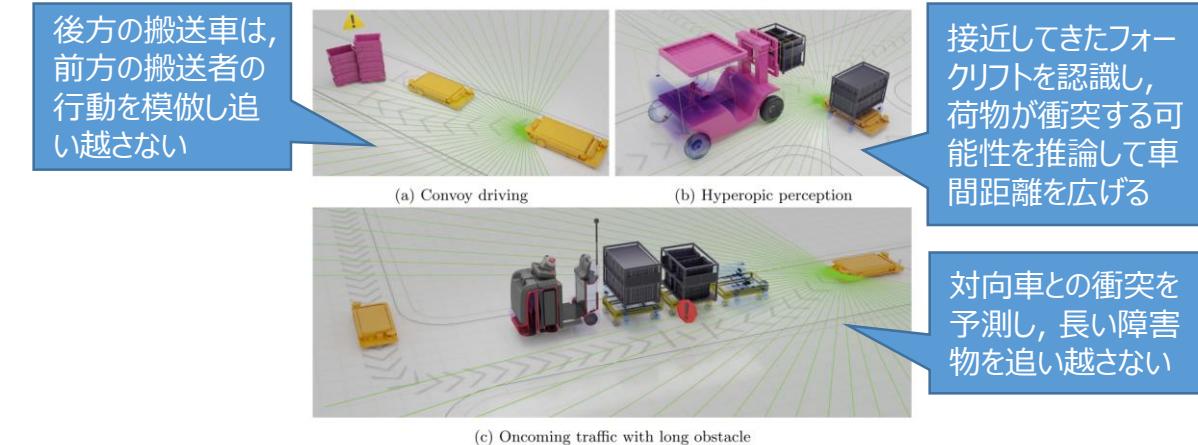


Fig. 1. Pictures of impediments encountered in a case study on autonomous robots deployed in an automobile manufacturing plant.  
ケーススタディから得られた課題の例

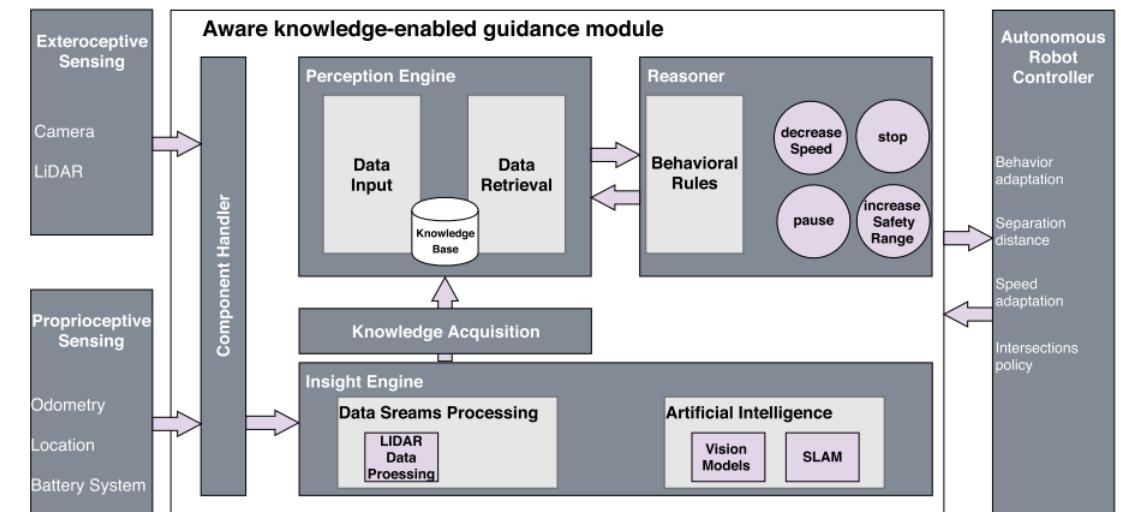


Fig. 3. An overview of the AWARE architecture.  
AWAREアーキテクチャ

# Semantic Integration of Bosch Manufacturing Data Using Virtual Knowledge Graphs

Free University of Bozen-Bolzano, Robert Bosch, Virtual Vehicle Research, Ontopic, University of Oslo, Umeå University

## ■ 概要

- 仮想知識グラフ（VKG）に基づくセマンティックデータの統合とアクセスを可能にするフレームワークSIB<sup>\*1</sup>をOntop上<sup>\*2</sup>に開発
  - \*1: Semantic Integration at Bosch
  - \*2: <https://ontop-vkg.org/>

## ■ 課題

- 製品の品質分析では、製造プロセスの各段階に設置された設備からデータの統合が不可欠であるが、クリーニング、重複排除、意味的な均質化などのデータ整備作業がコストになっている

## ■ 解決手段

1. SMT製造に関するドメイン知識を含むSMTオントロジーを構築
2. SMTオントロジーと装置データとをR2RMLによりマッピング
3. 製品品質分析タスクをSPARQLクエリのカタログとしてエンコード

## ■ 評価・実績

- 定性評価：分析タスクに必要なデータをSPARQLクエリで取得
- 定量評価：Apache SparkベースのSANSA-DLと性能比較

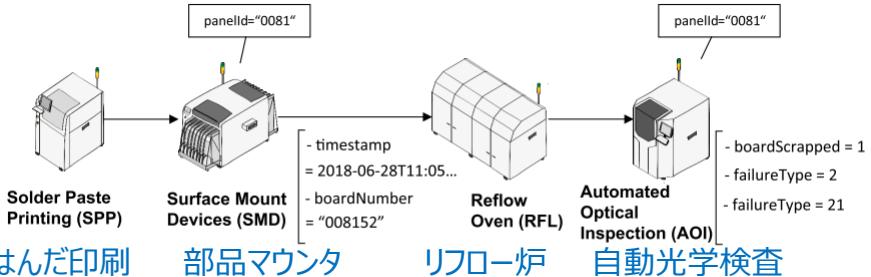


Fig. 2. Surface Mounting Process pipeline. The SMT process comprises four phases: SPP, SMD, RFL, and AOI. The machines, usually by different suppliers, rely on distinct formats and schemas for managing the same data across the process pipeline.

## Bosch工場における表面実装技術（SMT）のプロセス

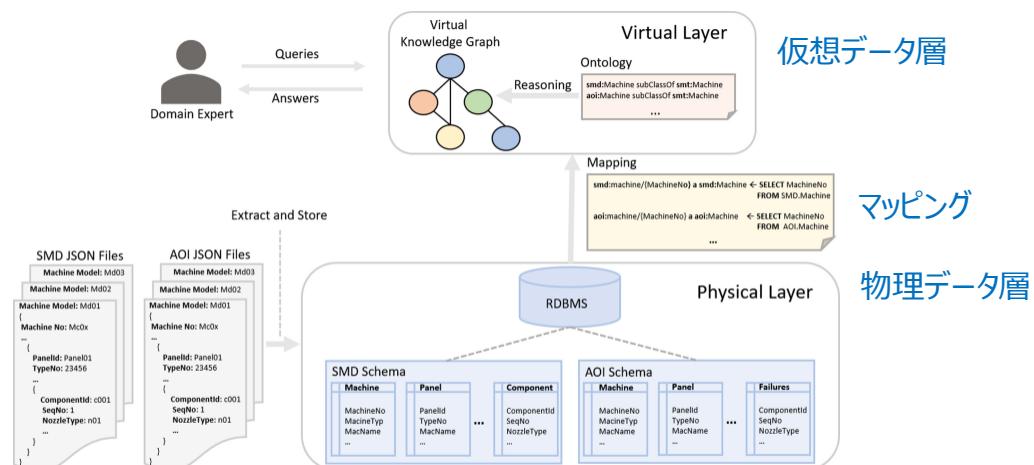


Fig. 1. Virtual Knowledge Graphs approach exemplified over Bosch SMT scenario.

Ontopを用いてBosch SMTシナリオで例示されたVKG手法

ご協力いただき  
ありがとうございました！