

ナレッジグラフ若手の会
ESWC2023サーベイ会

2023/07/19 オンライン

発表者一覧

※発表順です。サーベイ資料の公開時にセッション順に入れ替えます（青字部分）。

順	セッション	発表者
1	1A: KG Construction	eiichi_sunagawa
2	2A: KGs in the Real World I	cm3
3	1C: Alignment, Linking and Relation Extraction	eitiyamamoto
4	1B: Reasoning	福田研究室
5	2B: Question Answering and Knowledge Extraction	nomotom
6	3A: KGs and ML	森田研究室
7	3B: Completeness	古崎研究室
8	4A: SPARQL and Data Access	koji6_fujiwara
9	4B: KGs in the Real World II	acopom

順	セッション	担当者
10	5A: KG refinement and validation	古崎研究室
11	5B: KGs and Inductive Learning	森田研究室
12	Industry Track 1	kgyanagi
13	Industry Track 2	Nagano

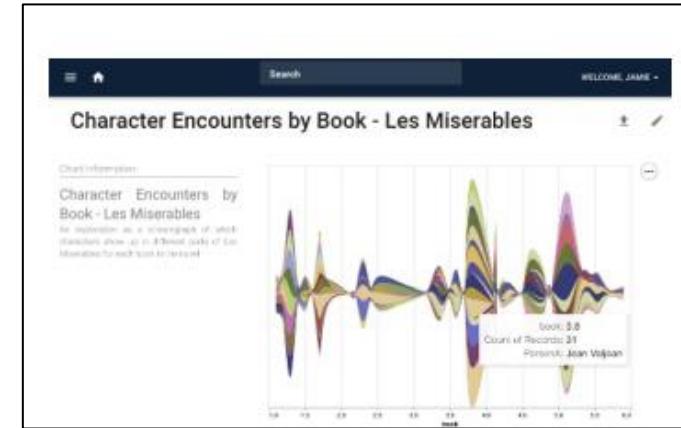
1A: KG Construction

eiichi_sunagawa

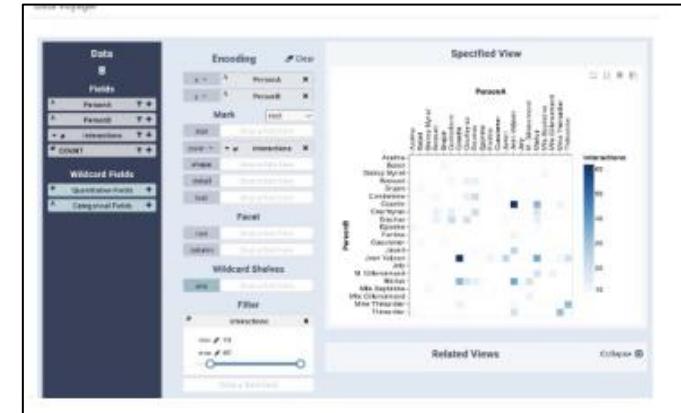
Whyis2: An Open Source Framework for Knowledge Graph Development and Research

Jamie McCusker and Deborah McGuinness
/ Rensselaer Polytechnic Institute, USA

- **一言でいうと**
KGのキュレーション、インタラクション、推論を備えたアプリ開発フレームワーク
- **Resource Type**
Software framework
- **動機**
KGを使ったアプリ開発に必要な技術を総合的に提供してくれるフレームワークがない
- **リソースの設計方針**
 - KG構築・維持のためのキュレーション、出所のトレース、矛盾発見の推論機能の提供
 - ナノスケールKGと呼ばれる小単位での管理を行う
- **新規性**
推論を使ったデータ取込み(移植)で他フレームワークより優れている
- **リソースの再現性／活用実績**
 - 材料科学や電波政策などの分野のKGアプリ開発で使われている
- **品質**
不明
- **リソースの可用性**
Pythonパッケージ、Docker-pullなどで利用可、Githubでソース公開
- **次にすべきこと**
チュートリアル等での利用者拡大、推論のベンチマーク、RDF Mapping Language対応



(a) レ・ミゼラブルの登場人物可視化

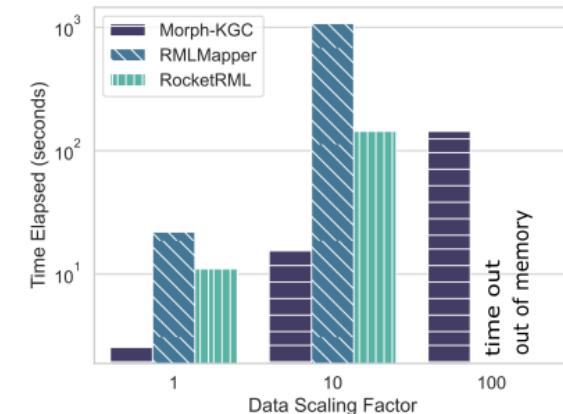


(b) (a)を開発するための
WhyisのData Voyager機能。

Boosting Knowledge Graph Generation from Tabular Data with RML

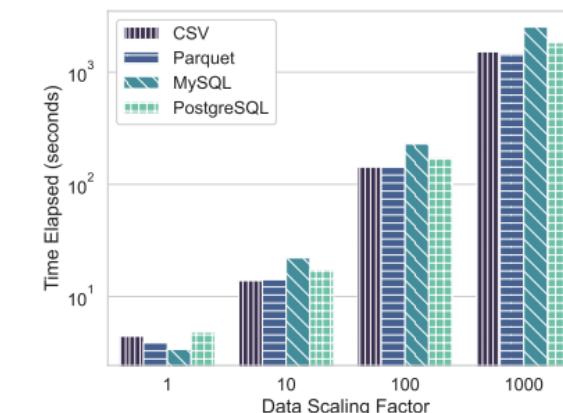
Julián Arenas-Guerrero, Ahmad AlObaid, María Navas-Loro, María S. Pérez and Oscar Corcho
 / Ontology Engineering Group, Universidad Politécnica de Madrid, Spain

- 一言でいうと
テーブルデータも扱えるようにRML Viewsを拡張した
- Resource Type
Software framework
- 動機
 - 表形式データをRMLで取り込む際に必要な前処理が手間
 - 拡張機能もあるがエンジニアに馴染みがなくスケールアップも困難
- リソースの設計方針
 - エンジニアが使いやすいようSQLクエリを組み込む
- 新規性
 - JSONデータやリストが入ったセルも扱える
 - FnOやFunULと違い、タームでなく論理ソースレベルで関数を定義できる
- リソースの再現性／活用実績
公共調達のシステムに適用されている
- 品質
GTFS-Madrid-Bench、LUBM4OBDAでベンチマーク
- リソースの可用性
Githubでソース公開
- 次にすべきこと



(a) GTFS-Madrid-Bench.

処理速度のベンチマーク。
Morph-KGC(提案技術、左の群青色バー)が高速。



(b) LUBM4OBDA benchmark.

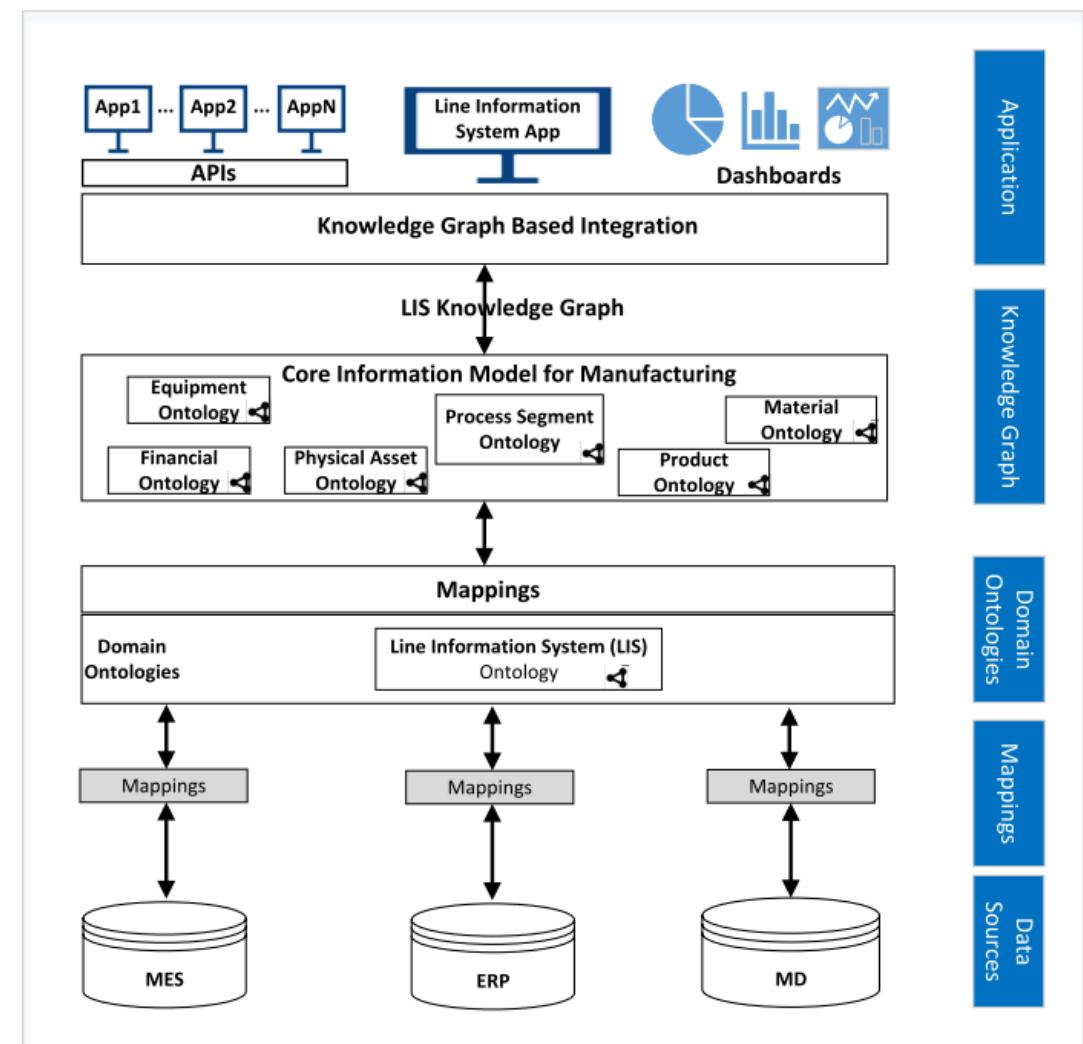
テーブル種類のベンチマーク。
異なるテーブル種類でも性能への影響が低い。

LIS: A Knowledge Graph-based Line Information System

Irlan Grangel-González¹, Marc Rickart², Oliver Rudolph², and Fasal Shah³

/ 1 Robert Bosch GmbH, Corporate Research, Germany 2 Robert Bosch GmbH, Automotive Electronics, Germany 3 Robert Bosch Engineering and Business Solutions Private Limited, India

- **一言でいうと**
生産ラインについての情報を統合的に扱えるようにした
- **動機**
生産ラインの情報を、様々なビジネス目的で利用できるようにしたい
- **手法**
KGを使ってデータを意味的に統合し、全方面での可視化、データの衝突回避、データ品質保証を提供した
- **背景**
生産ラインの情報が、製造実行(MES)、リソース計画(ERP)、マスタデータ(MD)といった複数のシステムでサイロ化されている
- **利用実績、結果**
 - Boschの工場12件で、1100以上のライン、16000以上の装置、400以上の工程に関するデータを統合
 - 経営者、ユーザ、開発者から肯定的なアンケート結果が得られた
- **学んだ教訓やベストプラクティス**
 - 生データを触っていたドメイン専門家が、必要な情報をクエリとして定式化できるようになった
 - データ品質をクリアにしたこと、MESのロケーション識別子が短すぎることに気付くなど改善につながった
 - KG技術のトレーニングをしたことで、抵抗していたドメイン専門家も納得できた

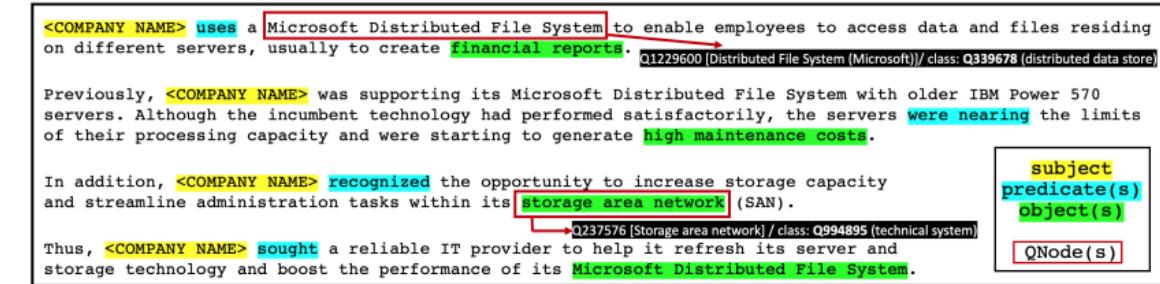


LISの全体アーキテクチャ。
下から、データソース／マッピング／オントロジー／知識グラフ／アプリの5層。

Understanding Customer Requirements – an Enterprise Knowledge Graph Approach

Basel Shbita¹, Anna Lisa Gentile², Pengyuan Li², Chad DeLuca², and Guang-Jie Ren²
 /¹ University of Southern California and Information Sciences Institute, USA /² IBM Research, USA

- **一言でいうと**
テキストから顧客要求を抽出して提供可能なソリューションを提案
- **動機**
顧客要求に企業が提供する製品・サービスをマッチさせたい
- **手法**
 - ビジネス要求文書、Open KG(Wikidata)、分野専門家の入力を合わせて事業KGを構築
 - 事業KGを用いて顧客要求をビジネスユニットに分類する
- **背景**
多くの場合は分野専門家が顧客要求と企業の製品・サービスをマッチさせる。しかし大企業の場合、資産、スキル、知財など見るべきものも複雑だし、顧客要求や製品仕様もダイナミック。また、IBMは膨大な業種を対象としており、提供技術も多用である。
- **結果**
過去10年以上蓄積した顧客要求と製品・サービスのデータセットを用いて構築したKGを用いて評価したところ、Doc2Vecだけを用いた分類より4.1%のF1スコア改善 (0.653→0.694) を確認
- **学んだ教訓やベストプラクティス**
技術説明が高度なアセットの場合については、製品と顧客の困りごと（「大量の注文の処理に苦労」など）の関係を示すことが理解に繋がった



文書からトリプルを抽出した例

<COMPANY NAME> -uses-> financial reports
 < COMPANY NAME> -were nearing-> high maintenance costs
 < COMPANY NAME> -recognized-> storage area network
 < COMPANY NAME> -sought-> Microsoft Distributed File System

2A: KGs in the Real World I

cm³

LauNuts: A Knowledge Graph to identify and compare geographic regions in the European Union

ヨーロッパの広域(NUTS)と狭域(LAU)の地域階層データをLOD化して、人口や面積や包含関係を検索できるようになりました。ツールも公開

Level	Minimum	Maximum	Example	Code 2021	Population
NUTS 0	Country level		France	FR	67.9 M
NUTS 1	3,000,000	7,000,000	Grand Est	FRF	5.5 M
NUTS 2	800,000	3,000,000	Alsace	FRF1	1.9 M
NUTS 3	150,000	800,000	Bas-Rhin	FRF11	1.1 M
LAU	-		Strasbourg	67482	0.3 M

NUTSでは人口範囲ごとに階層をつけているよ

<https://w3id.org/launuts>

Wikidataとかとの接続が
Future Work

Riccardo Tommasini, Filip Ilievski and Thilini Wijesiriwardene

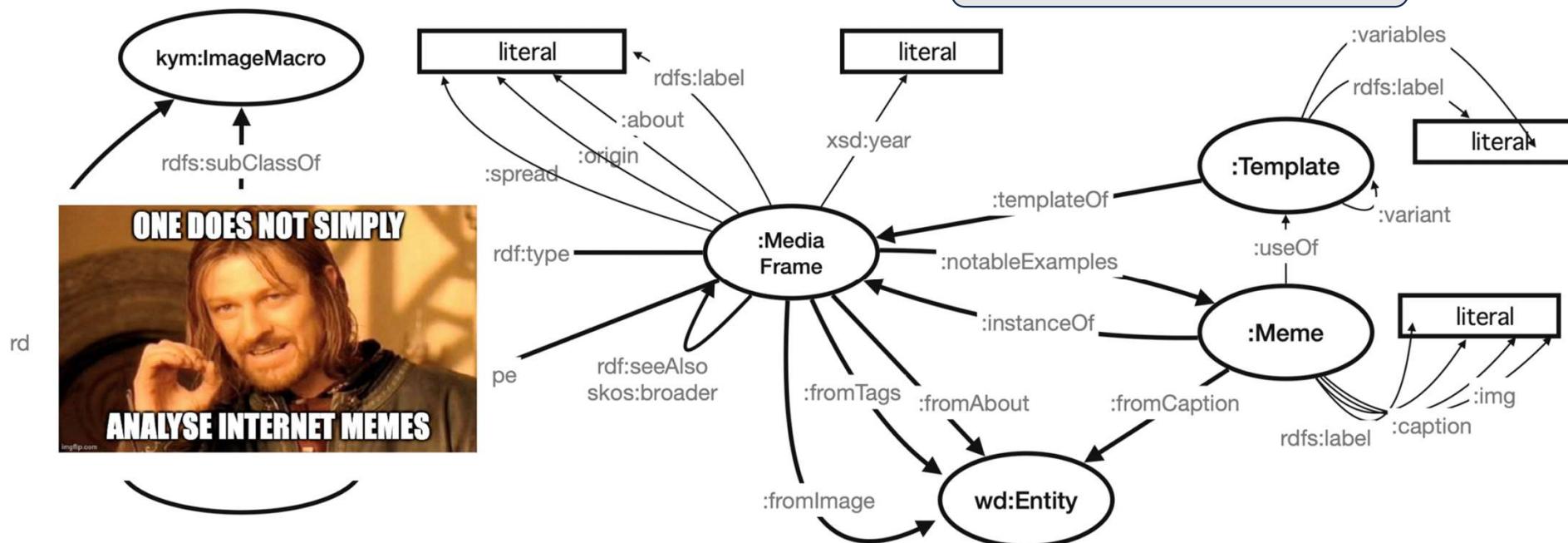
IMKG: The Internet Meme Knowledge Graph

Resource

100万以上のインターネットミームをRDFとかに整理して分析例も示してみた

ミームの特徴（マルチモーダル、簡約、内輪ウケ、変化する）から知識表現を設計。

<https://w3id.org/imkg>



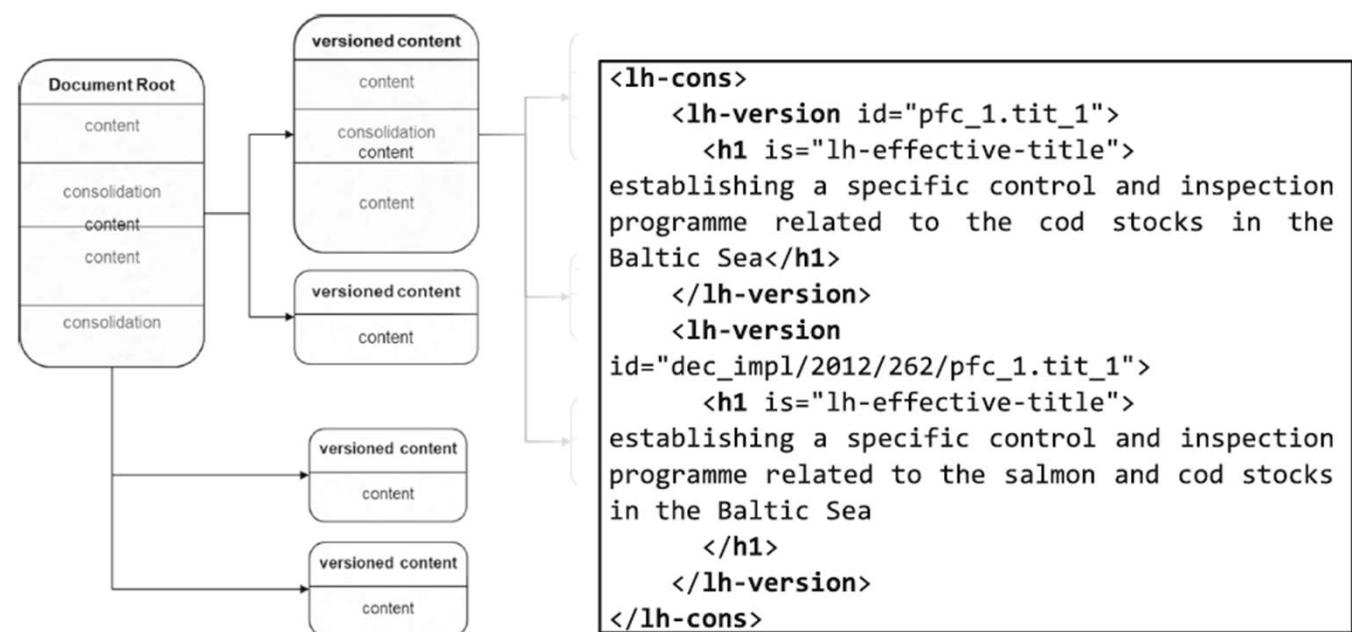
LegalHTML: A Representation Language for Legal Acts

欧洲委員会(EU)出版局が日々提供する法律行為のドキュメントをHTMLの拡張として提供するための技術開発を行っている、今年から実用予定

基本はXHTML的な発想。XMLベースでの可視化は不十分なのでHTMLを拡張する、今のHTML Living Standardならそれが可能だということ

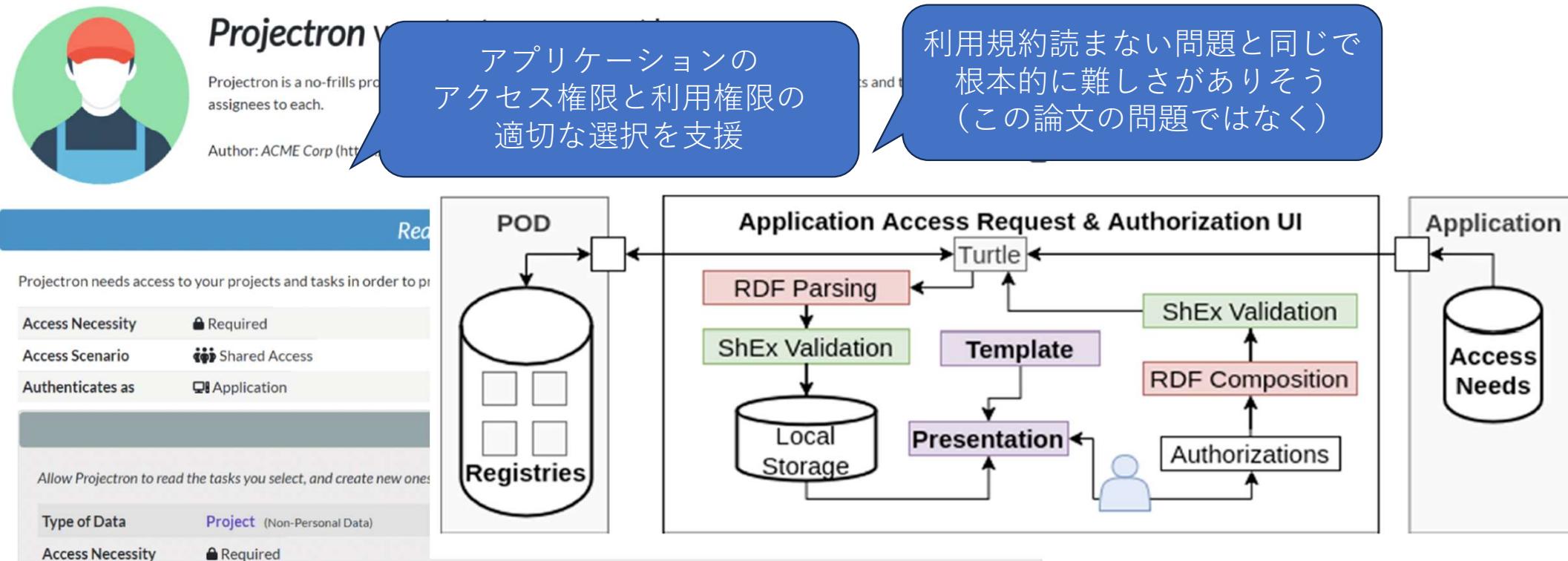
W3CのHTML5が
WHATWGのコレに
塗り替えられてた！

法律行為を構造化するだけ
でなく、加えられる変更を
表現するのに長けている



Prototyping an End-User User Interface for the Solid Application Interoperability Specification Under GDPR

GDPRで決められている個人情報コントロールのための設計(INTEROP Spec.)に対応したユーザインターフェースを作成して評価している



1C: Alignment, Linking and Relation Extraction

eitiyamamoto

Structural Bias in Knowledge Graphs for the Entity Alignment Task

Nikolaos Fanourakis, Vasilis Efthymiou, Vassilis Christophides, Dimitris Kotzinos, Evangelia Pitoura and Kostas Stefanidis

- 一言でいうと: 構造的多様性を考慮したデータセット作成した
- 動機: EAにおけるKGの構造的多様性に関する間接的なバイアスの影響を調査した
- 手法: 連結成分の数と個数、およびノードの次数を用いて定量化される
- 背景: 最先端のEA手法は、構造の特徴を利用している。しかし、間接的なバイアスを引き起こす可能性がある。
- 結果: 高いジャンプ確率でランダム・ウォークを用いたサンプリングは、元のKGに近いデータを作成した。
 - ジャンプ確率が高くなるほど結果が低くなる。エンティティ名を用いている手法は他に比べて影響されにくい。
- 考察: 確率論的手法は、構造の変化には強くない。埋め込み型は、利用する事実情報によって堅牢性が変わる。
- 課題: 構造的な多様性だけでなく、事実情報の多様性も考慮したサンプリング手法の拡張を目指している。

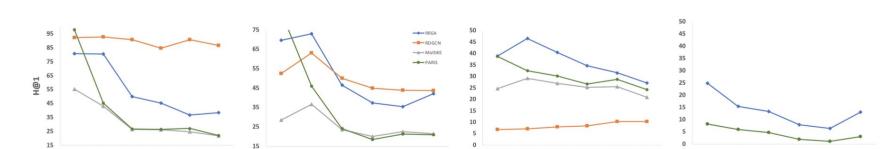
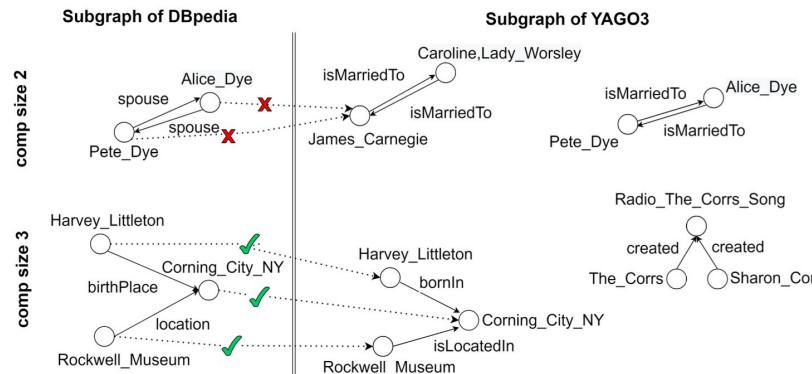


Fig. 3: Robustness of EA methods' effectiveness ($H@1$) to structural diversity brought by SUSIE sampling.

Transformer based Semantic Relation Typing for Knowledge Graph Integration

Sven Hertling and Heiko Paulheim

- 一言でいうと: トランスフォーマ・モデルを用いクラスの関係性を決める
- 動機: 2つのクラス間にどのタイプの関係が存在するかを決定する (eg. equivalent, sub/superclss, part of/has part)
- 手法: トランスフォーマ・モデルをファインチューニングし、異なるKGのクラスの関係を予測する。
- 背景: 現在のツールでは、同等のクラスしか特定できないため、複数のKGの統合には限界がある。
- 結果: 一番良い結果は、DistilBertをファインチューニングしたケースであった。
 - トレーニングではWordnet、DBpedia、Schema.orgを使用した。
- 考察: 'part of' や 'has part' は 'equivalent' や 'sub/superclass' よりも難しいことが分かった。
- 課題: 多言語の事例に対応し、可能な意味関係を増やす(例: is-a)

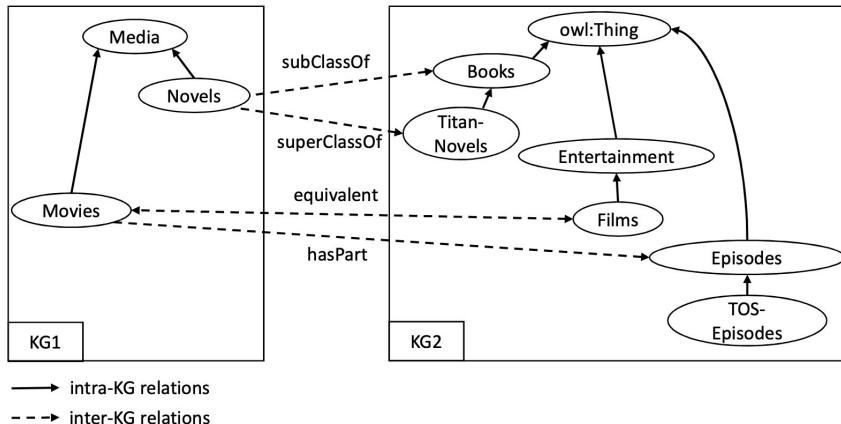


Table 5. Micro averaged F_1 results of all trained models differentiated by base model, datasets used for training, and kind of fine-tuning. STROMA serves as a baseline that is optimized for this gold standard. The top three values are printed in bold.

Base model	Dataset	Fine-tuning	G1	G2	G3	G4	G5	G6	G7
Albert	WDS	Task	0.422	0.673	0.906	0.125	0.314	0.261	0.651
	WDS	Test case	0.127	0.096	0.886	0.728	0.621	0.824	0.771
	WDS	+HP Tuning	0.434	0.746	0.912	0.721	0.663	0.761	0.940
	WDS	Test case +	0.540	0.758	0.886	0.213	0.361	0.211	0.566
	WDS	+Wikidata	0.091	0.096	0.895	0.757	0.645	0.810	0.807
	WDS	Test case +	0.575	0.797	0.878	0.779	0.710	0.817	0.687
	WDS	+HP Tuning	0.121	0.273	0.941	0.375	0.485	0.451	0.627
	WDS	Test case +	0.118	0.099	0.932	0.699	0.680	0.838	0.783
	WDS	+Wikidata	0.248	0.727	0.915	0.757	0.704	0.824	0.880
Bert	WDS	Task	0.534	0.752	0.919	0.228	0.385	0.261	0.639
	WDS	Test case	0.109	0.093	0.932	0.699	0.663	0.831	0.759
	WDS	+HP Tuning	0.348	0.718	0.927	0.779	0.726	0.852	0.723
	WDS	Test case +	0.560	0.741	0.896	0.250	0.385	0.268	0.651
	WDS	+Wikidata	0.103	0.087	0.913	0.699	0.663	0.817	0.735
	WDS	Test case +	0.440	0.758	0.904	0.691	0.716	0.845	0.747
	WDS	+HP Tuning	0.136	0.400	0.944	0.500	0.485	0.331	0.663
	WDS	Test case +	0.124	0.115	0.946	0.772	0.680	0.852	0.783
	WDS	+Wikidata	0.428	0.775	0.944	0.691	0.710	0.845	0.831
DistilBERT	WDS	Task	0.767	0.828	0.929	0.154	0.420	0.268	0.590
	WDS	Test case	0.100	0.085	0.883	0.699	0.698	0.852	0.783
	WDS	+HP Tuning	0.378	0.727	0.924	0.662	0.746	0.838	0.771
	WDS	Test case +	0.720	0.786	0.925	0.191	0.373	0.197	0.566
	WDS	+Wikidata	0.094	0.087	0.928	0.721	0.645	0.852	0.783
	WDS	Test case +	0.372	0.783	0.919	0.706	0.769	0.824	0.747
	WDS	+HP Tuning	0.183	0.510	0.946	0.331	0.367	0.317	0.675
	WDS	Test case +	0.115	0.144	0.942	0.728	0.704	0.866	0.771
	WDS	+Wikidata	0.419	0.735	0.934	0.743	0.734	0.838	0.771
Roberta	WDS	Task	0.799	0.820	0.916	0.272	0.373	0.204	0.554
	WDS	Test case	0.118	0.099	0.863	0.750	0.550	0.831	0.675
	WDS	+HP Tuning	0.363	0.445	0.936	0.735	0.627	0.810	0.759
	WDS	Test case +	0.799	0.845	0.841	0.250	0.296	0.155	0.494
	WDS	+Wikidata	0.124	0.099	0.883	0.728	0.669	0.803	0.663
	WDS	Test case +	0.472	0.727	0.898	0.728	0.669	0.824	0.747
	WDS	+HP Tuning	0.021	0.090	0.886	0.287	0.479	0.275	0.518
	WDS	Test case +	0.115	0.101	0.883	0.728	0.680	0.838	0.771
	WDS	+Wikidata	0.268	0.811	0.882	0.743	0.740	0.831	0.771
STROMA	-	-	0.761	0.792	0.854	0.765	0.716	0.866	0.807

NASTyLinker: NIL-Aware Scalable Transformer-Based Entity Linker

Nicolas Heist and Heiko Paulheim

- 一言でいうと: NIL エンティティと対応する言及クラスタを認識する EL アプローチ
- 動機: ELの主な課題: テキスト内のエンティティの曖昧性と、対応するエンティティが知識ベースにない言及 (NIL)
- 手法: NILエンティティの言及と言及 - エンティティの関連性を基に、言及とエンティティのクラスタを生成する。
- 背景: 適応データセットを使用する方法もあるが、 NILエンティティの言及をすべて含む必要があり、容易ではない。
- 結果: KGの作成タスクにおいて、既知のエンティティに確実にリンクし、 NILエンティティを特定できることが分かった。
- 考察: 既知のエンティティのクラスタの正しさは、 NILエンティティよりも低い。
 - NILエンティティの平均的なメンション数が少ないため、既知のクラスタでは 1つのメンションが欠落する可能性が高くなる。
- 課題: エンド・ツー・エンドのパイプラインを作成し、パフォーマンスを低下させることなく実行時間を短縮するための言及の検出を追加して、埋め込み作成時に NILエンティティを考慮する学習手順を含める。

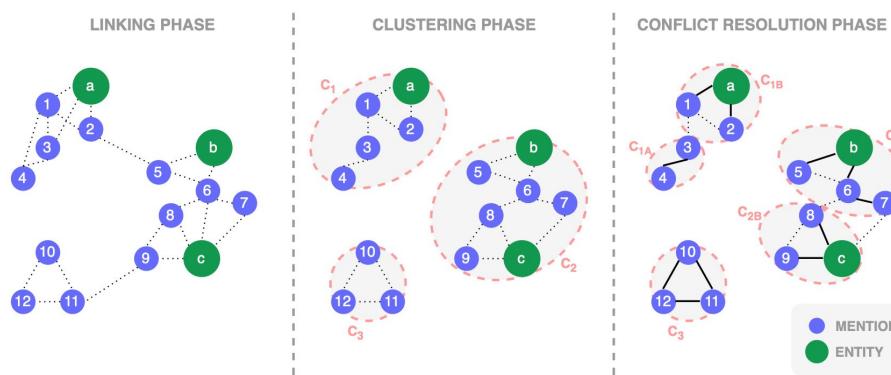


Table 2: Results for the test partition \mathcal{D}_{test}^N of the NILK dataset.

Approach	Known			NIL			Micro		
	F1	NMI	ARI	F1	NMI	ARI	F1	NMI	ARI
No Clustering	Exact Match	79.5	—	—	0.0	—	—	78.1	—
	Bi-Encoder	80.8	—	—	0.0	—	—	79.1	—
	Cross-Encoder	89.0	—	—	0.0	—	—	87.1	—
Clustering & No Reranking	Bottom-Up	64.6	99.0	97.5	41.6	94.8	81.8	64.1	96.8
	Majority	59.4	99.3	98.0	49.8	92.7	82.7	59.2	96.6
	NASTyLinker	76.8	98.6	95.3	40.8	95.2	76.8	76.0	97.3
Clustering & Mention Reranking	Bottom-Up	65.7	97.1	98.9	41.5	94.6	10.0	65.1	96.0
	Majority	66.6	92.4	74.8	44.0	94.4	73.2	66.1	92.4
	NASTyLinker	74.2	99.0	96.6	39.2	85.6	16.5	73.5	95.5
Clustering & Entity Reranking	Bottom-Up	89.0	99.3	96.2	41.6	94.1	58.0	87.9	98.2
	Majority	74.2	99.1	99.3	54.1	89.3	92.5	73.7	96.6
	NASTyLinker	90.4	99.3	95.5	43.7	94.6	85.3	89.4	98.5
Clustering & Full Reranking	Bottom-Up	84.2	99.6	98.9	41.8	84.6	3.2	83.3	96.2
	Majority	80.3	95.1	95.9	51.7	90.0	39.2	79.6	93.9
	NASTyLinker	87.9	99.5	99.2	42.5	87.6	33.6	86.9	97.4

RELD: A Knowledge Graph of Relation Extraction Datasets

Manzoor Ali, Muhammad Saleem, Diego Moussallem, Mohamed Ahmed Sherif and Axel-Cyrille Ngonga Ngomo

- 一言でいうと: オープンライセンスで公開している 8つの関係抽出データセットを一つの RDF知識グラフにまとめた
- 動機: データセットはさまざまな形で提供されていて、異なるタイプの関係抽出方法に関する実験を行うのが困難である。
- リソースの設計方針: 各関係と対応する文章／文書にはユニークな RDFリソースを当てて、
プロパティとして様々な統計／注釈が付加される。
- 新規性: 汎用的で、カスタムベンチマークやバランスの取れたデータセットを作成するために使用することができ、他の NLPタスクに使用することができます。
- リソースの再利用性/活用実績: オープンソースのコードは GitHubで公開されており、他のデータセットを変換するには簡単に拡張することができます。
- 品質: このモデルは、一般性、簡潔性、使いやすさ、互換性に従って作成された。
- リソースの可用性: ソースコード、データ、ドキュメントを含むホームページから一般公開されていて、 SPARQLエンドポイントも利用可能である。
- 次にすべきこと: 多言語と有料のデータセットを取り入れて、数関係と可変性の範囲を拡張する汎用的な関係抽出モデルを訓練する。

Table 2: Distribution of relations and sentences/documents in train, test and validation in different dataset. D represents documents, while M for a million

DATASET	Train		Validation		Test	
	relations	sent/docs	relations	sent/docs	relations	sent/docs
WEBNLG	246	74,779	186	72,719	246	80,710
NYT-FB	24	111,327	22	111,324	22	111,324
FewRel	64	44,800	16	11,200	0	0
SemEval	10	8,000	0	0	10	2,717
Google	4	16,948	0	0	0	0
WW	352	1,770,721	352	1,770,721	352	1,770,721
DocRed	96	3,053D	96	1000D	96	1000D
T-REx	685	6.02M/3M	0	0	0	0
Overall	1,481	8.04M/3M~	672	0.36M/1K	726	1.96M/1K

1B: Reasoning

福田研究室

A Comparative Study of Stream Reasoning Engines

Nathan Gruber and Birte Glimm/Ulm University, Germany

- 一言でいうと
ストリーム推論(SR)エンジンの比較研究.
- 動機
SRの標準的な比較手法が確立されていない.
- 手法
開発者に対する調査とベンチマークの2軸でSRエンジンを比較.
異なるベンチマークを跨げるインターフェースを提案(評価はなし).
- 背景
SRエンジンの正確さと比較可能性: RSP4J frameworkが既にある.
ベンチマーク: レイテンシやメモリ使用量を計測するCityBenchが既にある.
- 結果
開発者への聞き取り調査: 開発者は正確さや最大スループット, メンテナンス性などを重要視している.
ベンチマーク: RSP4J based SR の C-SPARQL2.0 と (厳密にはSRではないが) RDFox がよい成績を残した.
- 考察
正確さ(functional correctness)が実世界のアプリケーションにとって重要である. RSP4J がの高機能さが示された.
SRベンチマークのインターフェースを統一することは将来的なよりシンプルなベンチマークにつながる.
- 課題
RSP4J を拡張すること. SRエンジンの文脈では Java 以外のプログラミング言語も検討されるべき.
SRエンジンの完全性と信頼性を向上させる理論が求められている.
(理想的には定量化可能な形で) SRエンジンが入力に対して優先順位を付ける際の様子がクエリによらず分析できるとよい.

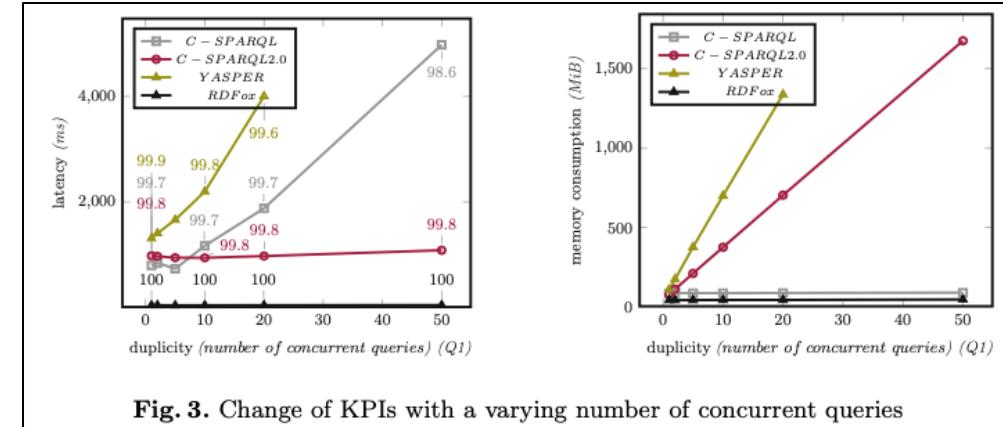


Fig. 3. Change of KPIs with a varying number of concurrent queries

Classifying sequences by combining context-free grammars and OWL ontologies

Nicolas L., Andrea P., Valentina, P. / University of Bologna, Italy

- **一言でいうと**
文脈自由文法をOWLで形式化するパターンと、それをシーケンス分類に利用する手法の提案
- **動機**
シーケンスを煩雑な操作をせずに情報を損なうことなく表現し、セマンティックwebの知識表現に基づくシーケンス分類を可能にすること
- **手法**
記述論理に基づいて文脈自由文法に新しい定義を与えることで、文脈自由文法を形式化するパターンを定義
文脈自由文法解析とOWLオントロジーを組み合わせシーケンス分類を行う
- **背景**
形式文法は様々な分野で有効活用されている
形式文法の一種である文脈自由文法はシンボルの意味情報が曖昧
曖昧性を軽減できる手法は存在するが情報が損失する可能性や、人間が理解できる形式にするには追加のツールが必要
- **結果**
オントロジーと従来の解析アルゴリズムを組み合わせることで、シーケンスを効果的に表現、分類することが可能ということを確認
- **考察**
提案手法は自然言語処理、生物情報学、プログラミング言語などの分野にも適用可能

A Concise Ontology to Support Research on Complex, Multimodal Clinical Reasoning

Sabbir M. Rashid, Jamie P. McCusker, Daniel Gruen, Oshani Seneviratne, and Deborah L. Rensselaer Polytechnic Institute, Troy, NY, USA

- ・一言でいうと
より簡潔な臨床の推論オントロジーの提案
- ・Resource Type
Ontology
- ・動機
臨床医が推論タスクをこなすのをより良くまねて多様な推論の方法を支援する必要がある。
- ・リソースの設計方針
既存のガイドラインと、糖尿病オントロジーのデザインの基準に従って設計した。
- ・新規性
既存の臨床の推論オントロジーは、大きくて複雑であり多くの時間を消費する可能性がある。今回はより簡潔なオントロジーである。
- ・リソースの再現性/活用実績
論文の中でドキュメントやgithubへのリンクが公開されている。
- ・品質
CBKを発展させた三つのガイドラインの原則に基づいて評価した。
- ・リソースの可用性
<https://github.com/tetherless-world/diabetes-pharmacologyontology>
- ・次にすべきこと
クエリをより早く実行する為に、期待される概念を含む説明のみを返すこと。

10 S. Rashid et al.

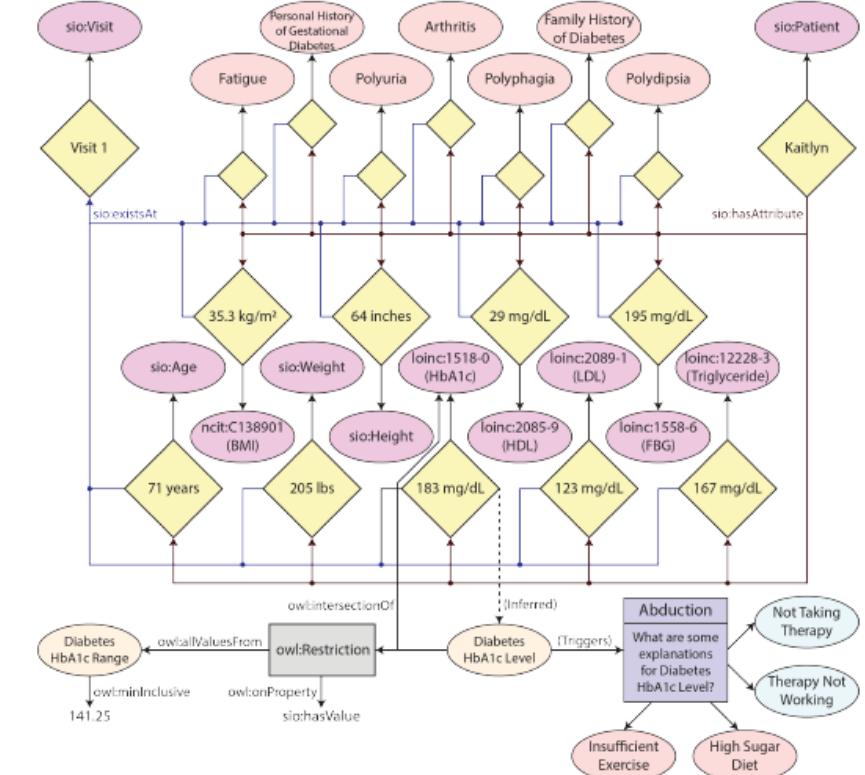


Fig. 2. A Personal Health Knowledge Graph includes instances of patients. A top-level ontology like SIO along with concepts in DPO is used for representing patient attributes. Deductive rules from the ontology are used to infer attribute categorizations. Ontology classes are shown as ellipses. External classes are pink while internal classes are color-coded as described in Figure 1. Instances are shown as yellow diamonds.

SemReasoner - A high-performance Knowledge Graph Store and rule-based Reasoner

Kevin Angele^{1,3}, Jürgen Angele², Umutcan Simsek¹, and Dieter Fensel^{1,2} Semantic Technology Institute Innsbruck, University of Innsbruck, Technikerstrasse 21a, 6020 Innsbruck, Austria, ²adesso, Competence Center Artificial Intelligence, ³Onlim GmbH, Weintraubengasse 22, 1020 Vienna, Austria

- 一言でいうと
包括的で拡張性を持ち、高い知識グラフストア及びルールベースの推論エンジンをもつSemReasonerの提案
- 動機、背景
異なるデータベースからのデータの統合時、不完全な知識が生まれてしまう。
この問題を解消し知識グラフにおける効率的な推論とクエリ応答をおこないたい
- 手法
 - グラフストアと演繹的な推論エンジンを組み合わせて、オントロジーやドキュメントを検索できるようにした。
 - モジュールアーキテクチャによって既存のITランドスケープやアプリに拡張性を持たせた。
 - OpenRuleBenchからテストケースを用いて他の最新のルールベースの推論エンジンと比較評価
- 結果
 - negation test caseにおいては最も遅かった。
 - Join 1 と Datalog Recursionにおいては最も早い
- 利用実績
Adesso SEという会社のいくつかの製品で使われている、
ルールによる意思決定や取引制限に違反する条件のリアルタイムでの監視などに貢献
- 学んだ教訓やベストプラクティス
SemReasonerは大量のデータを処理する能力があり、多くのテストケースにおいて他の推論エンジンより優れた結果を示した。

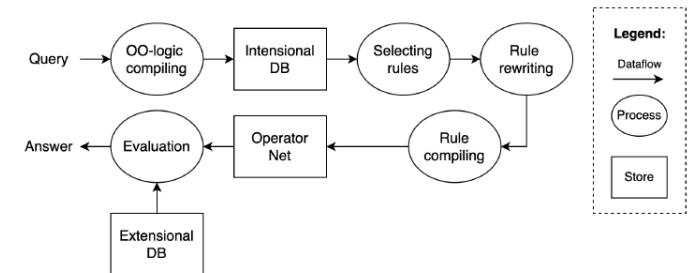


Fig. 2. Query processing in SemReasoner

2B: Question Answering and Knowledge Extraction

nomotom

Entity Linking for KGQA Using AMR Graphs

Nadine Steinmetz/Technische Universität Ilmenau, Bergische Universität Wuppertal

- 一言でいうと
KGQA(question answering on KG)のエンティティリンクで、質問テキストを抽象的意味表現(AMR)のグラフに変換し、エンティティの表層形認定に利用する手法の紹介
- 動機 記載なし
- 手法
 - 質問をAMR(抽象的意味表現)のグラフ構造に変換し、固有表現の表層形(nameノードの子ノードのラベル)を得る。
 - 事前学習モデルに問題(大文字/小文字の区別, 記号("’"))があり、自動生成した拡張データで学習をやり直した。言語モデルのノードラベル置き換えにも対応。
 - 抽出した表層形をKGのエンティティの辞書にマッピング。
 - 辞書情報: エンティティのラベル表記の情報(代替ラベル, 小文字正規化, ステミング, 主ラベルとの距離), 文脈情報(エンティティのクラスのラベル)等。
 - エンティティ候補を以下のスコアでランキング : ラベルベースのスコア, 入次数(入辺の数(ページリンク等)), 文脈スコア
- 背景 不明(他の手法を挙げているが、問題の記載なし)
- 結果
KGQAのテストデータ(QALD-9, LC-QuaD2.0)による評価で既存手法を上回った。
- 考察
学習データの拡張と再学習により、recallとprecisionを少なくとも5%改善できた。
- 課題
 - 今後の予定
 - 学習データの追加 (先頭大文字の語連続では表層形を分割しないよう学習、等)
 - AMRグラフの質問中の句が言語モデルに置き換えられないよう、学習、解析
 - パラメタを検証
 - 他のDBpedia、Wikidataベースのデータセットでの評価

Table 3. Results of our approach compared to other approaches

	QALD-9			LC-QuaD 2.0		
	R	P	F ₁	R	P	F ₁
DBpedia Spotlight	73.1	70.1	71.6	52.5	23.3	30.8
TagMe	70.2	52.2	59.9	59.4	29.5	37.4
AIDA	n.a.	n.a.	n.a.	30.5	38.5	33.1
BLINK (amrllib built-in)	64.4	61.0	62.5	n.a.	n.a.	n.a.
BLINK (standalone)	79.0	74.6	76.7	n.a.	n.a.	n.a.
OpenTapioca	n.a.	n.a.	n.a.	42.0	29.0	35.0
Falcon 2.0 API	78.0	68.1	72.7	56.0	50.0	53.0
Our approach	89.1	85.8	87.4	65.3	70.3	67.7

GETT-QA: Graph Embedding based T2T Transformer for Knowledge Graph Question Answering

Debayan Banerjee, Pranav Ajit Nair, Ricardo Usbeck and Chris Biemann/Universität Hamburg, Indian Institute of Technology (BHU)

- 一言でいうと

知識グラフを対象としたQuestion Answering (KGQA)で、質問テキストからSPARQL クエリを生成する際、Text-to-text(T2T)の事前学習言語モデル(PLM)の出力したラベルや埋め込みを用いて、KGのノードにグラウンディングしたクエリを生成する手法の紹介。

- 動機

事前言語学習モデル(PLM) T5の言語処理能力をKGQAの意味解析に利用したい。エンティティの識別子の生成方法を学習させ、グラウンディング、曖昧性解消に利用したい。

- 手法

- T5で質問からラベルと切り詰めた(truncated)KG埋め込みを含むクエリの原型を生成
- T5のデコード負荷削減のため埋め込みを切り詰める(ベクトルの次元数削減、floatの桁数)。エンティティ識別子には曖昧性解消に無力なIDの代わりにラベルを利用。
- エンティティと関係のラベルをKGノードにグラウンディングしたSPARQLクエリを生成
- エンティティラベルはラベルの検索結果を候補として埋め込みの類似度でリランク。関係ラベルはBERT埋め込みベースでWikidataのプロパティとマッチング。

- 背景

- ニューラル機械翻訳、PLMによるSPARQLクエリ生成でラベルや埋め込みを生成した例はない。PLMでKG埋め込みを学習した例はあるが、埋め込みを出力したものはない。

- 結果

LC-QuAD2.0、SimpleQuestions-Wikidataで手法(特に切り詰めた埋め込み)の有効性を確認。

- 考察

T5にKG埋め込みを学習させることで、KGQAのパフォーマンスが改善する場合がある。

- 課題

埋め込みの学習に特化した損失関数で埋め込みを生成。識別子は埋め込み以外も探す。

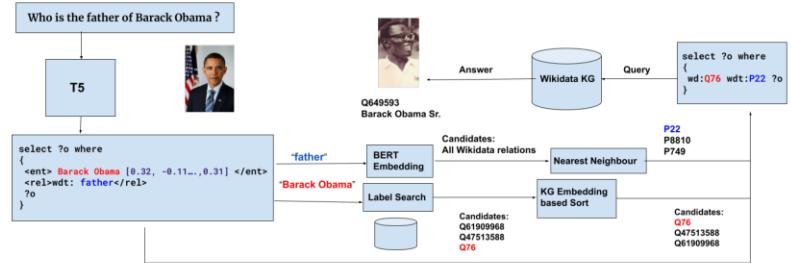


Fig. 1: Architecture of GETT-QA: T5 generates a skeleton SPARQL query with entity and relation labels and a truncated KG embedding string. The entity labels are grounded using label based search and embedding based sorting, while the relations are grounded using BERT embedding based sorting. The final query is executed against a KG to fetch the answer.

	P@1
UNIK-QA	0.005
Pullnet	0.011
Platypus	0.036
QAnswer	0.308
UNIQORN	0.331
GETT-QA without truncated embeddings	0.327 ± 0.002
GETT-QA (with truncated embeddings)	0.403 ± 0.0

Table 3: Results on LC-QuAD 2.0

	F1
KEQA	0.405
Text2Graph	0.618
GETT-QA without truncated embeddings	0.752 ± 0.004
GETT-QA (with truncated embeddings)	0.761 ± 0.002

Table 4: Results on SimpleQuestions-Wikidata

K-Hub: a modular ontology to support document retrieval and knowledge extraction in Industry 5.0

Anisa Rula, Gloria Re Calegari, Antonia Azzini, Davide Bucci, Alessio Carenini, Ilaria Baroni, Irene Celino/University of Brescia, Cefriel – Politecnico di Milano

- 一言でいうと
作業者の操作に役立つ知識のハブとなるよう、製造に関する知識の様々な側面をモジュール化したオントロジの概念モデルの紹介
- Resource Type Ontology
- 動機
製造現場のメンテナンスプロセスを支援する音声アシスタント開発を目指す共同研究・技術革新プロジェクト(Manufacturing Knowledge Hub)で、二社の従業員の情報アクセスを支援する音声アシスタントを開発した。文書検索、知識抽出に利用されるK-Hubオントロジはそのコアとなる。
- リソースの設計方針
文書のアノテーション、検索、知識抽出のカバー率を高めるため、複数の観点のモジュール(文書からの知識抽出、分野のトピック、プロセス情報、企業固有の情報)で構成された、拡張性の高いオントロジのハブを生成。他のシナリオにも利用できるよう、製造業分野の専門家の要求を一般化。
- 新規性
製造業の分野でオントロジをモジュール化し知識のハブを生成する初めての取り組みであり、同分野での標準化、コミュニティ主導の取り組みに役立つことを目指している。
- リソースの再現性／活用実績
- (「リソースの可用性」を参照。) 上記音声アシスタントで利用。Industry Ontology Foundry (IOF) Initiative に登録予定(2023/7/15時点で未確認)。
- 品質
- オントロジの評価はLinked Open Terms(LOT)プロセスにしたがい、ステークホルダである二社のメンバからフィードバックを受けた。
- オントロジの実装とcompetency questions及びfactsのリストのコンプライアンスチェックも実施。
- リソースの可用性
- オープンライセンスで無償利用可。企業固有のモジュール以外は自由に拡張可能。他のオントロジに追加モジュールとしてplug-in可能。
- 公開モジュールとプライベートモジュールでは公開方法を変え、GitHubリポジトリのアクセス設定が異なる。
- 次にすべきこと
- 機械学習/深層学習を知識抽出に利用等

Activity Recommendation for Business Process Modeling with Pre-trained Language Models

Diana Sola, Han van der Aa, Christian Meilicke, Heiner Stuckenschmidt/SAP Signavio/Data and Web Science Group, University of Mannheim

- 一言でいうと
ビジネスプロセスのアクティビティ推薦(新しいアクティビティのラベリング)において、
転移学習を用いて学習データに含まれないモデルやアクティビティに対応する手法の提案
- 動機
アクティビティの推薦(例: notify about outcome)にプロセスモデルのリポジトリから抽出
した分野特有の知識を利用していたが、汎用的な知識を追加して、推薦可能なアクティビティを拡張したい。
- 手法
プロセスモデルの形式意味論とラベルの意味論を用いて、アクティビティ推薦を複数のテキストのsequence-to-sequenceタスクとして扱えるよ
うにし、事前学習済みのsequence-to-sequence言語モデルT5を利用。
- 背景
 - プロセスモデルのリポジトリから抽出した知識を利用する手法では、未知のアクティビティから構成されているプロセスモデルに対応できず、
新しいアクティビティラベルを推薦できない。
 - ラベルの意味論に基づく手法には転移学習を利用するものもあったが、fine-tuningなしの事前学習で、推薦可能なアクティビティの範囲が学習
用データに存在するものに限られていた。
- 結果
 - プロセスモデルのコレクションによる実験で推薦の意味的なaccuracyが従来手法より高いことを確認(Table3: BLEU@k, METEOR@k, cos@k)。
 - 未知のアクティビティで構成されるプロセスモデルに対応可能で、学習データの語彙にないラベルの推薦が行えることを示した。
- 考察
NLPの転移学習をアクティビティ推薦に最大限に利用した最初の手法である。
- 課題
 - 手法の制限: ノードシーケンス長の制約、言語化の拡張(タスクタイプや組織の観点からみた情報の取り込み)、(実際の)推薦の有用性の調査
 - ルールベース手法(Sola et al.(2022)[47])と今回の手法のアンサンブル

List size	Approach	Hits@k	BLEU@k	METEOR@k	cos@k
$k = 10$	Sola et al. [47]	0.3102	0.3358	0.4149	0.5925
	BPART5	0.2800	0.3876	0.5154	0.6679
$k = 1$	Sola et al. [47]	0.0625	0.0714	0.1049	0.2539
	BPART5	0.0322	0.1179	0.2269	0.4112

Table 3: Results of the evaluated approaches

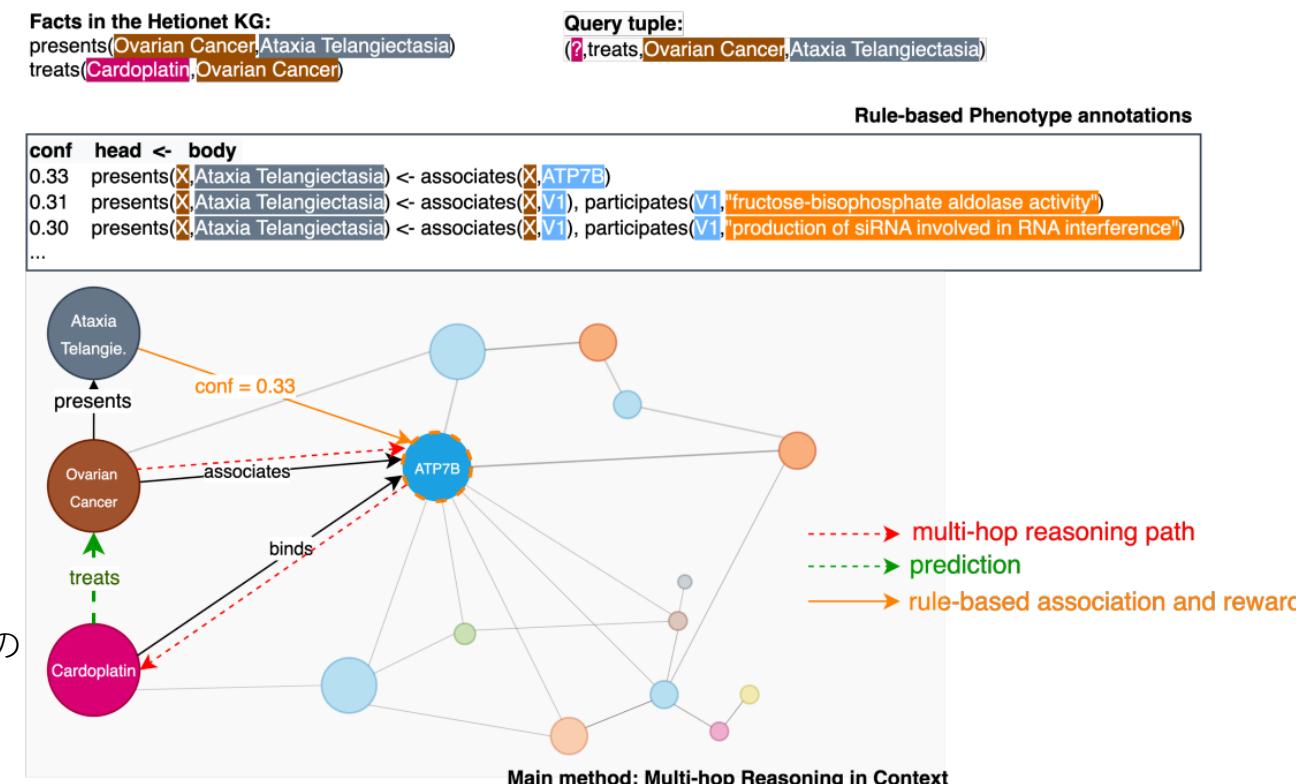
3A: KGs and ML

森田研究室

Explainable Phenotype-Centric Drug Repurposing via Deep Reinforcement Learning

Lise Stork, Ilaria Tiddi, René Spijker and Annette ten Teije

- 一言でいうと
薬剤再利用の説明機構の検索を改善する手法を提案
- 動機
薬剤の作用機序を理解し、新しい薬剤のターゲットを発見するために、薬剤再利用の解釈性メカニズムを向上
- 手法
深層強化学習と文脈整合に基づくマルチホップ推論手法、論理ルールマイニング、グラフ理論を組み合わせている
- 背景
強化学習ベースのマルチホップ推論は薬の再利用において新リンクの説明力を持つパスの発見が主な課題である
- 結果
実験結果から、この手法は潜在的な薬剤再利用の発見において高い精度と解釈性を持っていることが示された
- 考察
論理規則マイニングとグラフ理論の組み合わせが薬の再利用における推論パス発見に効果的であると結果を考察したが、説明の明確性には改善の余地があると指摘した
- 課題
注意機構やグラフニューラルネットワークの利用、他領域への適用性、生物学的注釈の充実、専門家の関与を今後の研究方向と提案した



Describing and Organizing Semantic Web and Machine Learning Systems in the SWeMLS-KG

Fajar J. Ekaputra, Majlinda Llugiqi, Marta Sabou, Andreas Ekelhart / Vienna University of Economics and Business

- 一言でいうと
既存のSWeMLS（セマンティックウェブと機械学習の双方に依存するシステム）を機械的に実行可能な形で記述したKGであるSWeMLS-KGの提案
- Resource Type
Knowledge Graph
- 動機
セマンティックウェブを機械学習へ活用することへの需要の高まり
- リソースの設計方針
SWeMLSに関する記述のある約500の論文からデータを収集、選択、抽出
- 新規性
SWeMLSを機械的に実行可能な形で記述した初のオントロジー
- リソースの再現性／活用実績
Open Research Knowledge Graphによってインポートされ、容易にデータの閲覧、アクセス、引用、再利用することに貢献
- 品質
論文内でKGの作成で使用したクエリあり
- リソースの可用性
<https://semantic-systems.org/sites/swemls-kg/>
- 次にすべきこと
SWeMLSのテンプレート化

Combining Semantic Web and Machine Learning for Auditable Legal Key Element

Gerald Anna Breit¹, Laura Waltersdorfer², Fajar J. Ekaputra^{3,2}, Sotirios Karampatakis¹, Tomasz Miksa², and Gregor Kafer²

¹ Semantic Web Company, Vienna, Austria first.last@semantic-web.com

² TU Wien, Vienna, Austria first.last@tuwien.ac.at

³ Vienna University of Economics and Business, Vienna, Austria first.last@wu.ac.at

- 一言でいうと

現実世界のユースケースに基づいて、セマンティック Web と機械学習の手法を組み合わせて法的許可から重要な要素を抽出することを目的としたハイブリッド AI システムを開発および評価

- 動機

法的文書から重要な要素を抽出することは、間違いが発生しやすい複雑な作業

- 手法

セマンティック Web と機械学習を組み合わせ、ナレッジ グラフの形式で背景知識を組み込みながら、言語および埋め込みモデルを通じてデータを抽出

- 結果

定量的評価では全体的に有望な結果、詳細な定性分析では具体的なエラーの種類が明らかになり、現在のプロトタイプを改善する方法についての指針が得られた

- 利用実績

電子許可管理システム (EPMS) の運用、オーストリアでの公式許可に関する組織的および官僚的プロセス (申請、決定、修正など) を容易にする

- 学んだ教訓やベストプラクティス (不明)

Fig. 2: Overview over the developed Auditable Key Element Extraction System

休憩

3B: Completeness

古崎研究室

Evaluating Language Models for Knowledge Base Completion

Blerita Veseli/Max Planck Institute for Informatics , Sneha Singhania /Max Planck Institute for Informatics , Simon Razniewski/Bosch Center for AI , and Gerhard Weikum /Max Planck Institute for Informatics

- **一言でいうと**

知識ベース補完のための言語モデルで活用している従来のベンチマークより高性能なものを目指す研究。

- **動機**

現状の知識ベースは膨大な量の情報が格納されているが不完全である。言語モデルを用いることで、知識ベースのファクトを補完し、より完全な知識ベースを構築することができるようになる。

- **手法**

言語モデル（BERT）の学習に、WD-knownという自分たちが用意したデータセットを用いる。比較対象としてLAMA-T-REXを用いる。両方学習させた言語モデルも評価する。

- **背景**

既存のベンチマークでは、チューニング時に理想の結果になるようにしていって、実用性に欠けた精度が得られていた。

- **結果**

今回用意したデータセットでは言語モデルのファクト予測能力はLAMA-T-REXよりも高かった。

- **考察**

偏ったデータセットを用いるとファクト予測における言語モデルの性能が低くなる。それを改良するにはより大規模なデータセットを使用するべき。

- **課題**

もっと知識ベースを有効に活用するには、知識ベースに特化した言語モデルの構築が考えられる。

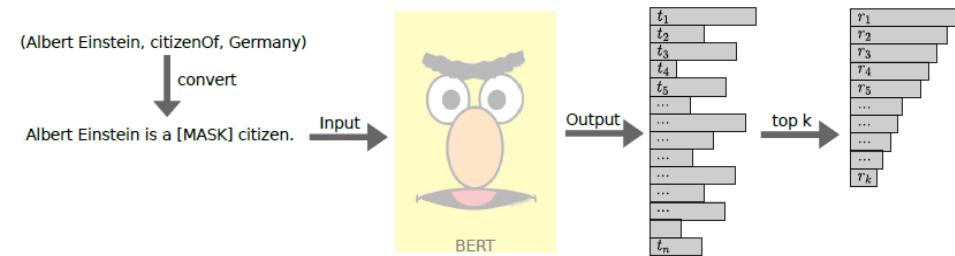


Fig. 1. To query the LM for an object slot, we convert triples into relation-specific prompts by masking the object, following [17]. The output is a probability vector over the model’s vocabulary. We retain the top $k = 10$ predictions.

Two-Fold Analysis

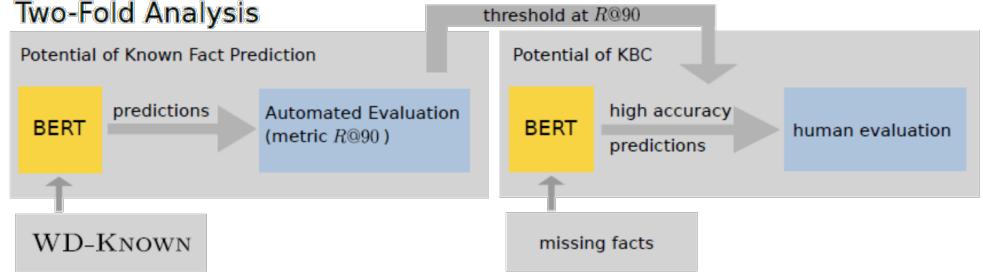


Fig. 2. Our systematic analysis is divided into two parts. First, we analyse existing-fact prediction by an automated evaluation computing the recall achieved at 90% precision ($R@P90$). The prediction at $R@P90$ is used as a threshold in the second part of the analysis. We analyse the potential of KBC for missing facts via a manual evaluation.

Dataset	#unique subjects	#unique objects	#triple	#multi-token objects	object dist. entropy
LAMA-T-REX	total	31,479	61,85	34,039	-
	average	767	150	830	0.76
WD-KNOWN	total	2,956,666	709,327	3,992,386	-
	average	72,113	17,300	97,375	0.67

Table 1. Our WD-KNOWN dataset in comparison to LAMA-T-REX. We report the total number of distinct objects (#unique objects), distinct subjects (#unique subjects), and the number of triples (#triples) as well as the total number of objects consisting of more than one token (#multi-token objects), and the average object entropy.

Subsumption Prediction for E-Commerce Taxonomies

Jingchuan Shi¹, Jiaoyan Chen¹, Hang Dong¹, Ishita Khan², Lizzie Liang², Qunzhi Zhou², Zhe Wu², and Ian Horrocks¹

¹ Department of Computer Science, University of Oxford, UK, ² eBay Inc., USA

- **一言でいうと**
電子商取引におけるタクソノミー内のサブサンプション予測
 - **動機**
電子商取引上の検索と在庫管理の促進
 - **手法**
学習されたBERTを使用しインスタンス埋め込み
 - **背景**
商品サイトにおいてユーザーの検索結果のカテゴリは売上やレビュー影響する
 - **結果**
ボックス埋め込みベース予測はK-NN予測よりも時間と空間を大幅に節約
 - **考察**
既存のクラスから新しいクラスを推論し分類を充実させることを示す
 - **課題**
カテゴリーが階層構造の異なる枝に属しているとサブカテゴリーと見なされない

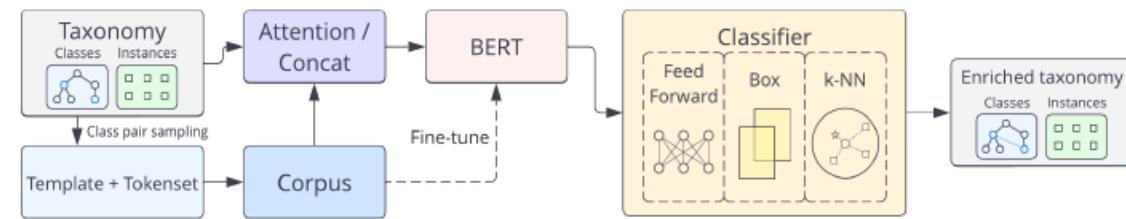


Fig. 2: Framework overview: (i) corpus construction, (ii) fine-tuning, (iii) joint embedding of label paths and instances, (iv) subsumption prediction using feed forward layer / box embedding / extensional inference with k-NN

Table 2: Results of predicting masked subsumptions of the eBay taxonomy

Table 3: Results of predicting masked subsumptions of the Alibaba taxonomy

Two-view Graph Neural Networks for Knowledge Graph Completion

Vinh Tong/University of Stuttgart, Dai Quoc Nguyen/Oracle Labs, Dinh Phung/Monash University, Dat Quoc Nguyen/VinAI Research

- 一言でいうと
知識グラフCompletionのためのニューラルネットワークモデルWGEの提案.
- 動機 Completion向の新しいGNNモデルを提案したい.
- 手法
知識グラフから**single undirected entity-focused graph**（エンティティをノードとしたグラフ）および**single undirected relation-focused graph**（述語をノードとしたグラフ）という**2つのView**によるグラフを生成し、これをエンベディングして、エンコーダ/デコーダモデルによるGNNで学習する.
- 背景
知識グラフは様々な応用に用いられているが、知識グラフには完備でない部分が含まれるため、それを補完する技術の研究が盛んに行われている.
- 結果
7つのベンチマークを対象とした比較実験 (Table.2) により提案手法が既存手法より良い結果 (Hits@10で1-2%) を得た.
- 考察
アブレーション解析により述語をエンティティ化したグラフが、関係間の潜在的な依存関係に良い影響を与えていることなどが示された.
- 課題 記載なし

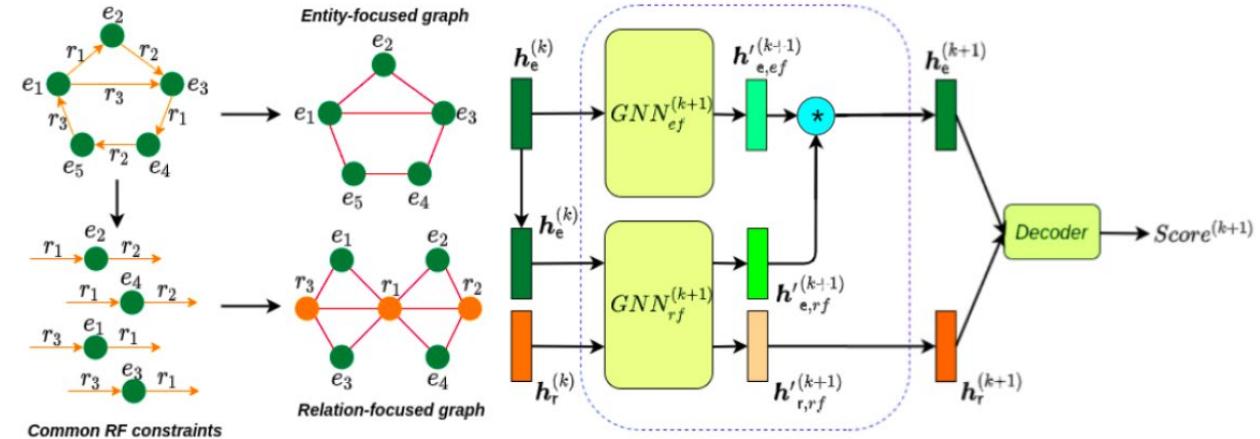


Fig. 1. An illustration of our proposed WGE.

Method	CoDEx-S	CoDEx-M	CoDEx-L	LitWD1K	LitWD19K	LitWD48K	FB15K-237	
	MRR H@10	MRR H@10	MRR H@10	MRR H@10	MRR H@10	MRR H@10	MRR H@10	
TransE	0.354	63.4	0.303	45.4	0.187	31.7	0.313	51.3
ComplEx	0.465	64.6	<u>0.337</u>	47.6	0.294	40.0	0.413	67.3
ConvE	0.444	63.5	0.318	46.4	0.303	42.0	0.477	71.4
TuckER	0.444	63.8	0.328	45.8	0.309	43.0	0.498	74.4
R-GCN	0.275	53.3	0.124	24.1	0.073	14.2	0.244	46.2
CompGCN	0.395	62.1	0.312	45.7	0.304	42.8	0.323	52.8
SimQGNN	0.435	<u>65.2</u>	0.323	47.7	0.310	43.7	<u>0.518</u>	<u>75.1</u>
QuatE	0.449	64.4	0.323	<u>48.0</u>	<u>0.312</u>	<u>44.3</u>	0.514	73.1
WGE	<u>0.452</u>	66.4	<u>0.338</u>	48.5	<u>0.320</u>	<u>44.5</u>	<u>0.527</u>	76.2

Table 2. Experimental results on seven test sets. Hits@10 (H@10) is reported in %.

4A: SPARQL and Data Access

koji6_fujiwara

Join Ordering of SPARQL Property Path Queries

Julien Aimonier-Davat, Hala Skaf-Molli, Pascal Molli, Minh-Hoang Dang, and Brice Nédelec/LS2N, University of Nantes, France

- 一言でいうと
プロパティパスクエリ(PPQ)の評価コストを把握できるコストモデルと動的計画法アルゴリズムを提案する。

- 動機
PPQは任意の長さのパスをマッチングを含み、評価が難しい。

- 手法
 $C2RPQ_{UF}$ と呼ばれるUNIONとFILTERを含む形式のクエリに対し、既存のエンジンを変更せずに効率的な順序を見つける。

- 背景
PPQはWikidataで頻繁に使用されており、クエリログの38%を占めるが、その多くはタイムアウトする。

- 結果
最新のWikidata Query Benchmarkにおいて提案手法がVirtuosoやBlazeGraphよりも有意に優れた結合順序を見つけることを示した。

- 考察
より優れたモデルを利用することで最適化時間を短縮できる。
推定量の信頼区間を計算することでランダムウォークを系統的にではなく、各結合順序にコストを適用できるかもしれない。

- 課題
提案手法の入れ子スター型への一般化。
MINUS、OPTIONAL、FILTER NOT EXISTSが存在する場合のランダムウォークについての検討。

```
SELECT DISTINCT ?x1 ?x3 WHERE {
?x1 wdt:P641 wd:Q3609 . # tp3
?x1 wdt:P17 ?x3 . # tp2
?x3 wdt:P361 wd:Q27611 . # tp1
}
```

(a) $(Q_1^{J_2})^{1..1}$

```
SELECT DISTINCT ?x1 ?x3 WHERE {{
?x1 wdt:P641 wd:Q3609 . # tp3
?x1 wdt:P17 ?x3 . # tp2
?x3 wdt:P361 wd:Q27611 . # tp1
} UNION {
?v1 wdt:P641 wd:Q3609 . # tp3.1
?v1 wdt:P641 ?v1 . # tp3.2
?x1 wdt:P17 ?x3 . # tp2
?x3 wdt:P361 wd:Q27611 . # tp1
}}
```

(b) $(Q_1^{J_2})^{1..2}$

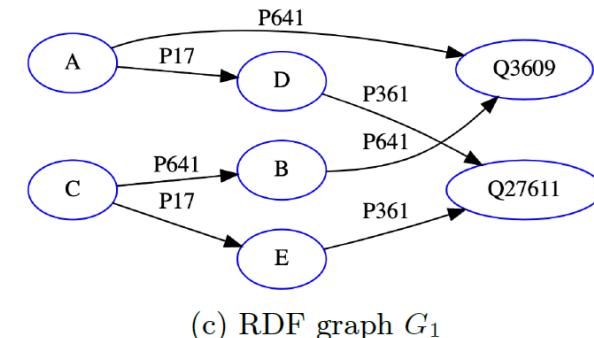


Fig. 3: RDF graph G_1 and rewrites of the query $Q_1^{J_2}$ used to estimate the cost of J_2 with random walks.

Enabling Live SPARQL Queries Over ConceptNet Using Triple Pattern Fragments

Marcelo Machado¹, Guilherme Lima¹, Elton Soares¹, Rosario Uceda-Sosa², and Renato Cerqueira¹

¹IBM Research Brazil, Rio de Janeiro, Brazil, ²IBM TJ Watson Research Center, Yorktown Heights, NY, USA

- 一言でいうと

トリプルパターン・フラグメント(TPF)インターフェースと
知識グラフフレームワークComunicaを用いて、
ConceptNetに対するライブSPARQLクエリを可能にした。

- 動機

大規模な推論用知識グラフConceptNetにSPARQLクエリを
オンラインで実行したい。

- 手法

RDFグラフのインターフェースであるTPFを通す。

- 背景

ConceptNetはSPARQLに対応していない。

ダンプを取ってオフラインで変換して実行する方法もあるが、すぐに試すことができない。

- 結果

SPARQLクエリをConceptNet上で評価することができた。

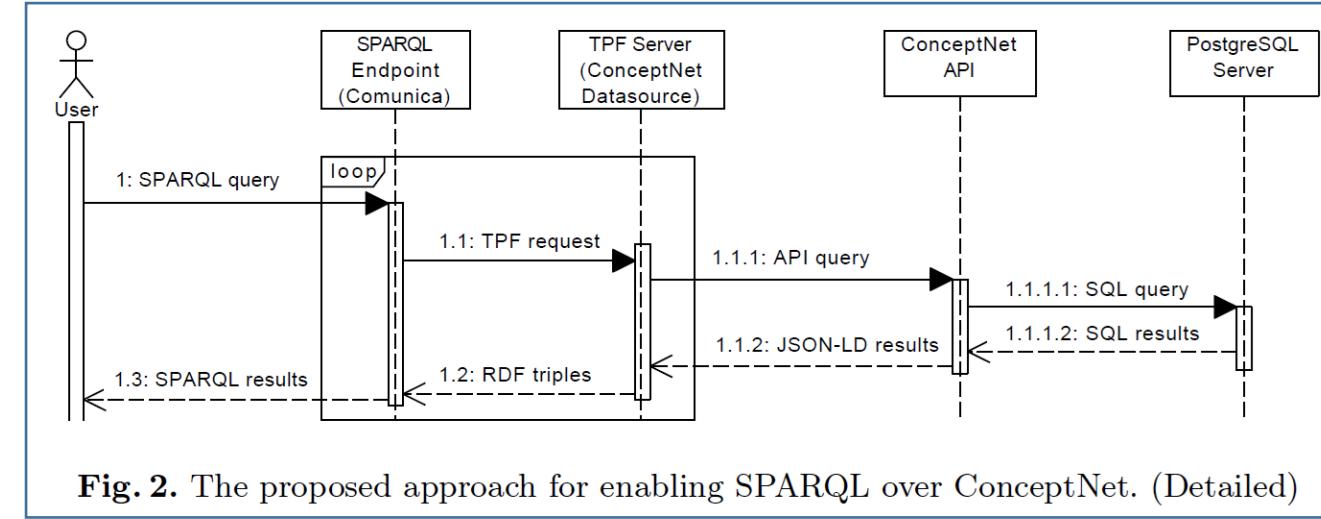
発生するオーバーヘッドは小さく、ConceptNetの内部エッジ表現を最適化することでさらに削減できることが示された。

- 利用実績

<https://github.com/IBM/tpf-conceptnet-datasource>

- 学んだ教訓やベストプラクティス

将来的にはTPFを通さずにComunicaとConceptNetを直接接続したい。



Evaluation of a Representative Selection of SPARQL Query Engines using Wikidata

An Ngoc Lam, Brian Elvesæter and Francisco Martin-Recuerda/SINTEF AS, Norway

- 一言でいうと

6つのツールに対してWikidataの328のユーザー定義クエリを使って性能評価を行った。

- 動機

Wikidataは著名なソフトウェア成果物となりつつあるが、サイズが大きく、効率的な処理ツールが求められている

- 手法

各ツールに対し、約163億トリプルのWikidataを読み、クエリを実行した結果について6つの指標で評価する。

- 背景

既存のベンチマークとして、合成データセット9つと実世界データセット6つに言及している。

- 結果

RDFoxはImportが高速であり、Exportが成功した唯一のツールだった。

クエリ実行もRDFoxが総合的に優れていて、次点でGraphDB。GraphDBとStardogはマシンメモリに依存しない結果を示した。

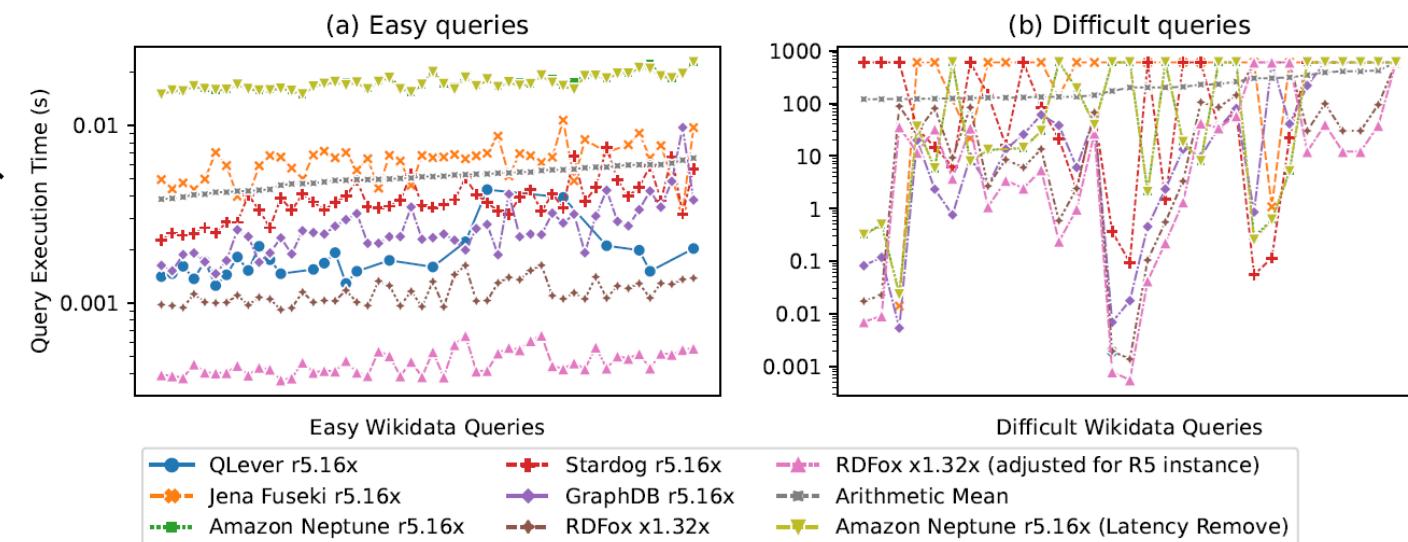
- 利用実績

記載なし

- 学んだ教訓やベストプラクティス

一部ツールにおいてクエリ結果に不一致が出たが、その原因是不明だった。

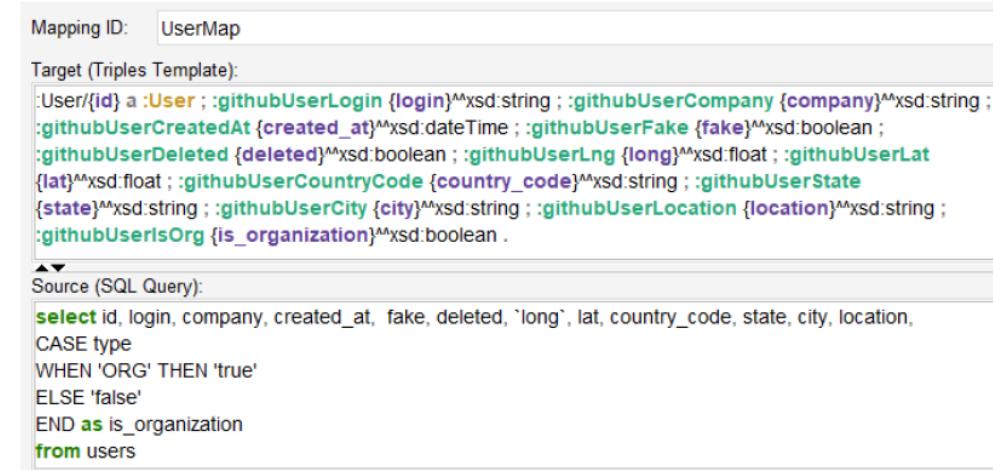
今後はより多くのトリプルストアのコレクションの評価、Wikidataの評価で使ったクエリをWDBenchを使って拡張する予定である。



Investigating Ontology-based data access with GitHub

Yahlieel Jafta^{1,3}, Louise Leenen^{1,3}, and Thomas Meyer^{2,3/1} University of the Western Cape, Cape Town, South Africa
² University of Cape Town, Cape Town, South Africa, ³ Centre for Artificial Intelligence Research (CAIR), South Africa

- 一言でいうと
OBDA(Ontology-based data access)によってGitHubのデータ利用を強化する。
- 動機
データ規模が大きくなるほど、そのアクセスがボトルネックとなっている。
- 手法
OBDAツールのOntopを使い、GitHubのミラーであるGHTorrentデータセットとGitのドメインオントロジーであるSemanGitを拡張して、OBDAを実装する。
- 背景
OBDAでは、データソースをオントロジーにマッピングして仮想化し、SPARQLクエリを実行することができる。
- 結果
OBDAのドメイン語彙によるクエリは、オントロジーの公理に対する推論と組み合わされ、有望な結果を示した。
- 利用実績
将来的に、拡張オントロジーを公開し、インターフェースとAPIエンドポイントを公開することで、GitHubコミュニティの評価に貢献する
- 学んだ教訓やベストプラクティス
異種データからの知識抽出の分野における人工知能のより広い領域を促進する



The screenshot shows a mapping interface with the following details:

Mapping ID: UserMap

Target (Triples Template):

```
:User/{id} a :User ; :githubUserLogin {login}^^xsd:string ; :githubUserCompany {company}^^xsd:string ;
:githubUserCreatedAt {created_at}^^xsd:dateTime ; :githubUserFake {fake}^^xsd:boolean ;
:githubUserDeleted {deleted}^^xsd:boolean ; :githubUserLng {long}^^xsd:float ; :githubUserLat
{lat}^^xsd:float ; :githubUserCountryCode {country_code}^^xsd:string ; :githubUserState
{state}^^xsd:string ; :githubUserCity {city}^^xsd:string ; :githubUserLocation {location}^^xsd:string ;
:githubUserIsOrg {is_organization}^^xsd:boolean .
```

Source (SQL Query):

```
select id, login, company, created_at, fake, deleted, 'long', lat, country_code, state, city, location,
CASE type
WHEN 'ORG' THEN 'true'
ELSE 'false'
END as is_organization
from users
```

Fig. 2. Mapping assertions for the User entity



The screenshot shows a mapping interface with the following details:

Mapping ID: CommitMap

Target (Triples Template):

```
:Commit/{id} a :Commit ; :commitSha {sha}^^xsd:string ; :commitAuthor :User/{author_id} ;
:committedBy :User/{committer_id} ; :commitCreatedAt {created_at}^^xsd:dateTime .
```

Source (SQL Query):

```
select id, sha, author_id, committer_id, created_at from commits
```

Fig. 3. Mapping assertions for the Commit entity

4B: KGs in the Real World II

acopom

KGs in the Real World II

担当: 東京都市大学 山口敦子

An Upper Ontology for Modern Science Branches and Related Entities

Said Fathalla, Christoph Lange and Sören Auer

A knowledge graph of contentious terminology for inclusive representation
of cultural heritage

Andrei Nesterov, Laura Hollink, Marieke van Erp and Jacco van Ossenbruggen

MOSAIK: An Agent-Based Decentralized Control System with Stigmergy For
A Transportation Scenario

Sebastian Schmid, Daniel Schraudner and Andreas Harth

HHT : an approach for representing temporally-evolving historical territories

William Charles, Nathalie Hernandez and Nathalie Aussénac-Gilles

An Upper Ontology for Modern Science Branches and Related Entities (Resource Track)

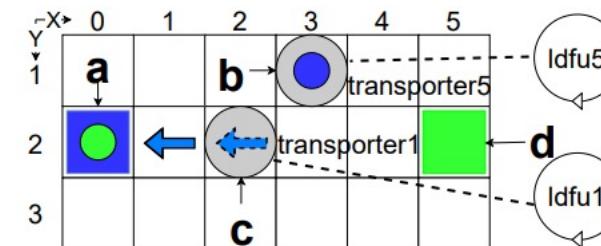
- 一言でいうと
様々なドメインの科学知識を統一的に扱うための上位オントロジーModern Science Ontology (ModSci) を提案
- Resource Type: Ontology
- 動機
主要な科学分野から新興の科学分野まで、様々な分野の科学知識を整理する、実用的なセマンティックモデルを開発したい。
- リソースの設計方針
リソースの設計方針は、以下の5段階から成る **The Systematic Approach for Building Ontologies (SABiO)**に従っている。 (1) オントロジーキャプチャー (2) 形式化 (3) 設計 (4) 実装 (5) 評価。構築は反復的に行われ、主要な分野から構築を始めて、分野を広げていく方針を取った。また、よく知られたオントロジー(SWEET, SKOS, FOAF, BFOのRoleなど)とのマッピングを行った。
- 新規性
学術文献の分類に利用でき、それに基づいて、研究者に論文推薦ができる。学際的な分野の索引に利用できる。研究トピックに基づく、研究者や組織の分類ができる。
- リソースの再利用性/活用実績
GitHubに公開することで再利用性を高めた。また、人間が読むために、Widoco wizard(オントロジー用のドキュメンテーションツール)でHTMLドキュメンテーションを作成した。
- 品質
SABiOのガイドラインに基づき、データに基づく評価と形式面の評価を行った。データに基づく評価はテストデータを作成し、利用シナリオに沿った質問にModSciを使って答えられるか検証した。形式面はオントロジーが正しく構築されていることを確認した。
- リソースの可用性
作成したリソースは<https://github.com/saidfathalla/Science-knowledge-graph-ontologies/tree/master/ModSci>にて、CC BY International ライセンスで利用可能。
- 次にすべきこと
ModSciをより洗練させること、カバーできる範囲を広げること、アプリケーションを作ること。

A knowledge graph of contentious terminology for inclusive representation of cultural heritage (best paper nominee) (Resource Track)

- 一言でいうと
英語とオランダ語の議論や問題がある用語を含む文化遺産のナレッジグラフ(LOD)を構築した.
- Resource Type: Knowledge graph
- 動機
文化遺産の説明に不適切な用語が含まれている場合がある。一方、それらの用語に関する専門知識は様々なリソースに分散しており、結果として用語の一貫性がない。
- リソースの設計方針
まず不適切な用語を収集して用語集を作成し、それを以下の原則に従ってナレッジグラフに変換した。
(1)元の説明の知識を保持している、
(2)既存のLOD語彙をできるだけ利用する、
(3)不適切用語問題に限らず広く利用が可能な設計とする。
- 新規性
(1)文化遺産の専門家のドメイン知識に基づいて用語集を作成したこと、
(2)文脈に応じて不適切な用語を見つけることができる、
(3)用語はWikidataやWordNetなどのLODリソースのエンティティとリンクしていること。
- リソースの再利用性/活用実績
リソースの設計方針の(2)(3)はリソースの再利用性を高めるためのものである。
- 品質
不適切になりうる英語の用語(75)、オランダ語の用語(83)はドメイン専門家によって利用方法、提案、代替用語とリンクされている。
- リソースの可用性
<https://w3id.org/culco#>にて、CC-BY-SA 4.0で公開している。
- 次にすべきこと
不適切な用語に対して代替の用語推薦を利用する。

MOSAIK: An Agent-Based Decentralized Control System with Stigmergy For A Transportation Scenario (In-Use Track)

- 一言でいうと
　　スティグマジーという分散制御手法を輸送制御に応用するフレームワークMOSAIKの提案
- 動機
　　輸送制御にスケーラビリティ, 適応性, 柔軟性をもつ分散制御手法を使いたい.
- 手法
　　スケーラビリティの向上のため, 内部状態を持たないエージェントを利用する. エージェント間のコミュニケーションは共通の環境を使用することで間接的に行う(スティグマジー). 本研究では共通の環境として, RESTful Read-Write Linked Data interfaceを用いる.
- 背景
　　輸送制御にこれまで使われてきたマルチエージェントシステムなどの分散制御手法は, しばしば中央制御の一面が残っている. 分散システムの利点を失わない手法は自動運転などに役に立つと思われる.
- 結果
　　スティグマジーをLinked Data上で行うという提案手法を定式化したうえで実装し, Industry4.0を想定した一つの輸送制御の例を使って挙動を確認した.
- 学んだ教訓やベストプラクティス
　　スティグマジーとLinked Dataを組み合わせて自己組織化された分散制御システムを構築できた. 実用化に向けては, エージェント間の共有情報の最適な粒度, 衝突を避ける仕組みの検討, タスクによりどの範囲の情報まで取得する必要があるかが変わるので, その検討が必要.



輸送制御の例

HHT: an approach for representing temporally-evolving historical territories (Resource Track, best paper nominee)

- 一言でいうと
階層的な歴史的領土を表現するHHT(Hierarchical Historical Territory)オントロジーを構築した.
- Resource Type:Ontology, Software and datasets
- 動機
歴史的な領土の変遷を表すために地理幾何学的表現がしばしば用いられるが、過去の領土の概念は幾何学的表現はそぐわないことが多い。また、階層性を求められることもある。幾何学的表現を用いず、階層性を実現した領土の表現が欲しい。
- リソースの設計方針
TSN(Territorial Statistical Nomenclatures)オントロジーをベースとして、Nomenclaturesの代わりに領土の単位を表すhht:Unitクラスを導入した。さらに階層性を表す語彙を追加した
- 新規性
地理幾何学的表現を用いず、かつ、階層性を実現した、歴史的領土のオントロジーはこれまで存在しない。
- リソースの再利用性/活用実績
HHTを利用した領土の変化を検出するアルゴリズムを提案した。
- 品質
領土の変化検出アルゴリズムが、フランス第三共和政の領土変化など、いくつかの事例において、正しく検出できるか実験することで評価を行った。
- リソースの可用性
<https://zenodo.org/record/7451702>にて、Creative Commons Attribution 4.0 Internationalで利用可能。
- 次にすべきこと
時系列の表現。現在は時間ごとにオブジェクトが作られてしまい、グラフが大きくなってしまう。

5A: KG refinement and validation

古崎研究室

Refining Large Integrated Identity Graphs using the Unique Name Assumption

Shuai Wang/Vrije Universiteit Amsterdam, Joe Raad/LISN, University of Paris-Saclay, Peter Bloem/Vrije Universiteit Amsterdam, and Frank van Harmelen/Vrije Universiteit Amsterdam

- **一言でいうと**

大規模なアイデンティティグラフのためのUnique Name Assumption(UNA)の定義の提案。

- **動機**

既存のUNAより例外に対応でき、誤ったリンクを識別できるようにするUNAの定義。

- **手法**

それぞれ異なる定義のUNAをLODクラウドのデータで検証を行い、エラー検出の信頼性を評価する。

- **背景**

大規模な統合知識ベースのエラーの検出と修正を行うUNAは様々な条件で常には成立しないため、その精度を高める。

- **結果**

the internal UNA (iUNA)という新しいUNAを定義して他のUNAとの比較を行い、iUNAのほうが精度が高いという結果が出た。

- **考察**

他のUNAよりiUNAの精度が高いという結果は、例外を使用したモデリングが上手く行えているからと考えられる。

- **課題**

パラメータの自動調整や、ゴールドスタンダードの拡張を行う。

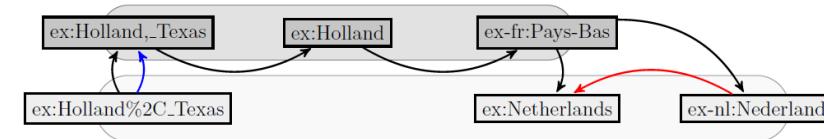


Fig. 2: An example CC with links expressing identity (black), redirection (red), and encoding equivalence (blue) (see also Section 3).

		Training set				Evaluation set			
		precision	recall	Ω	A	precision	recall	Ω	A
Louvain	res=0.01	0.020	0.803	0.091	39,471.4	0.042	0.727	0.087	42,424.2
	res=1.0	0.020	0.778	0.087	39,226.2	0.042	0.660	0.084	43,610.0
Leiden		0.249	0.198	0.377	3,398.4	0.068	0.323	0.439	2,782.6
MetaLink	t=0.9	0.076	0.029	0.522	241	0.086	0.032	0.524	337
	t=0.99	0.036	0.004	0.591	58	0.013	0.001	0.635	99
nUNA	label, w1	0.126	0.150	0.590	406.2	0.042	0.063	0.597	684.6
	label, w2	0.153	0.181	0.591	529.0	0.061	0.075	0.580	697.4
	comment, w1	0.201	0.146	0.595	263.0	0.098	0.040	0.618	356.4
	comment, w2	0.209	0.178	0.597	360.2	0.063	0.036	0.606	431.2
qUNA	w1	0.258	0.152	0.641	492.0	0.058	0.036	0.662	706.4
	w2	0.227	0.174	0.640	566.6	0.101	0.054	0.671	634.2
iUNA	label, w1	0.333	0.127	0.606	78.0	0.122	0.013	0.652	236.8
	label, w2	0.204	0.118	0.616	125.8	0.136	0.028	0.647	235.0
	comment, w1	0.267	0.090	0.598	63.8	0.097	0.002	0.636	141.2
	comment, w2	0.258	0.117	0.607	133.2	0.117	0.003	0.638	173.8

Table 4: Evaluation of the Louvain algorithm with two resolution values, the Leiden algorithm, MetaLink with two threshold values, and our algorithm using different UNA and settings.

Repairing EL Ontologies Using Weakening and Completing

Ying Li / Linköping University, The Swedish e-Science Research Centre

Patrick Lambrix / Linköping University, The Swedish e-Science Research Centre, University of Gävle

- 一言でいうと

ELオントロジの修復における、不要な公理を削除した場合の悪影響を軽減するための、弱体化と補完を組み合わせた対話的なアプローチの提案

- 動機

高品質なオントロジベースのアプリケーションを開発するためには、正しく完全性の高いオントロジーが重要

- 手法

削除と弱体化、補完を組み合わせるためのフレームワークを提案
弱体化と補完のための新しいアルゴリズムの提示

- 背景

不要な公理を削除することによるオントロジー修復では、オントロジーの領域で正しい結果を削除してしまうことがある

- 結果

13の組み合わせ戦略の比較

- 考察

正しさと完全性にはトレードオフが存在することを示唆

- 課題

弱体化と補完を統合した完全なデバッグについて検討する

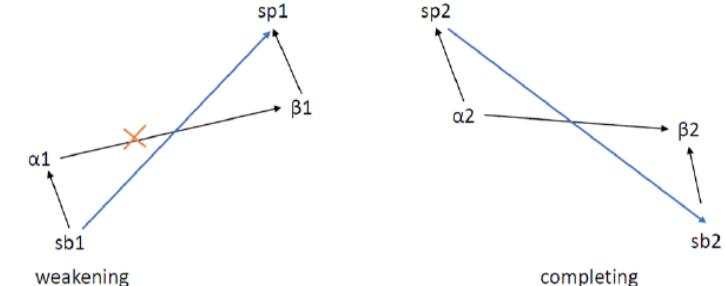


Fig. 1. Examples.

Table 2. Removing, weakening and completing - operations.

Operations	Description
R-all	Remove all the wrong axioms at once.
R-one	Remove the wrong axioms one at a time.
R-none	Remove nothing.
W-all	Weaken all wrong axioms at once.
W-one	Weaken the wrong axioms one at a time
C-all	Complete all weakened axioms at once.
C-one	Complete the weakened axioms one at a time.
AB-one	Add one wrong axiom back.
AB-all	Add all wrong axioms back.
AB-none	Add nothing back.
U-now	Update the changes immediately.
U-end_one	Update the changes after the iteration of each wrong axiom.
U-end_all	Update the changes after iterations of all wrong axioms.

A Framework to Include and Exploit Probabilistic Information in SHACL Validation Reports

Rémi Felin, Catherine Faron, and Andrea G. B. Tettamanzi/Université Côte d'Azur, Inria, I3S, Sophia-Antipolis, France

- 一言でいうと

SHACLによるRDFのValidationに確率的論的な情報を導入し, Validationの結果を真偽値(Boolean)のみでなく適切な割合で示せるように拡張した.

- 動機

SHACLによるValidationの結果を真偽値(Boolean)でなく適切な割合で得たい.

- 手法

RDFデータセットに含まれる制約違反のプロパティの割合を考慮した確率モデルに基づいたフレームワークを導入. 違反数を測定する尺度を定義. それらを用いてSHACLによるValidation結果のレポート内容を拡張した.

- 背景

The Shapes Constraint Language (SHACL) はRDFの制約を記述する言語として広く用いられている. しかし, 大規模なRDFデータにおいては, 一貫性がない部分が存在するのは不可避で, SHACLによるValidationでな実践的に不十分である.

- 結果

評価実験により既存のRDFデータセットのValidationを行った.

- 考察

拡張した内容により様々な観点で評価が可能. Validationの実行時間は, 標準的なSHACLによるもの(1分29秒)と比較し, さほど変わらなかった(1分35秒).

- 課題

より詳細なSHACLに対応する.

```

@prefix sh: <http://www.w3.org/ns/shacl#> .
@prefix psh: <http://ns.inria.fr/probabilistic-shacl/> .
@prefix : <http://www.example.com/myDataGraph#> .

# SHACL Standard
:v1 a sh:ValidationResult ;
  sh:focusNode :n1 ;
  [...]
  sh:sourceShape :s1 .

:v2 a sh:ValidationResult ;
  sh:focusNode :n2 ;
  [...]
  sh:sourceShape :s1 .

[...]

[ a sh:ValidationReport ;
  sh:conforms false ;
  sh:result :v1 ;
  sh:result :v2 ;
  [...] ]
# SHACL Extension
# shape s1
psh:summary [
  a psh:ValidationSummary ;
  psh:generality "0.2"^^xsd:decimal ;
  psh:numConfirmation 178 ;
  psh:numViolation 22 ;
  psh:likelihood "0.0806"^^xsd:decimal ;
  psh:referenceCardinality 200 ;
  psh:focusShape :s1
] ;
].

```

Fig. 2: Example of an extended SHACL validation report for a shape

iSummary: Workload-based, Personalized summaries for Knowledge Graphs

Giannis Vassiliou, Fanourias Alevizakis, Nikolaos Papadakis and Haridimos Kondylakis/Department of Electrical and Computer Engineering, HMU, Computer Science Department, UOC, Institute of Computer Science, FORTH

- **一言でいうと**

iSummaryという、個人の要求に応じて情報を要約する手法の提案。

- **動機**

情報を要約するという手法は様々だが、そのほとんどがユーザーのニーズや好みを取り込んでおらず、拡張性がない。

- **手法**

クエリログを活用することで、複数の重みの割り当てと計算問題を解決する方法と、クエリログの数に比例して要約の品質を保証するアルゴリズム。

- **背景**

ウェブ上で利用可能な知識グラフのサイズと複雑さが爆発的に増大したため。

- **結果**

3つのデータセットを使用して評価。実行時間の点 (Fig.6) では少し劣るが、カバレッジの点 (Fig.3) では比較対象全てを圧倒している。

- **考察**

既存の手法と比較してよい性能が認められた。

- **課題**

個人に合わせた要約が常に同じでないよう、多様性を導入することが考慮されている。

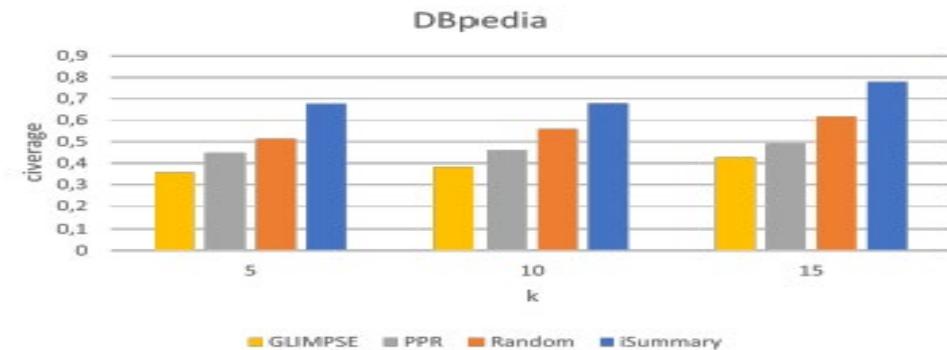


Fig. 3. DBpedia coverage for various k and baselines.

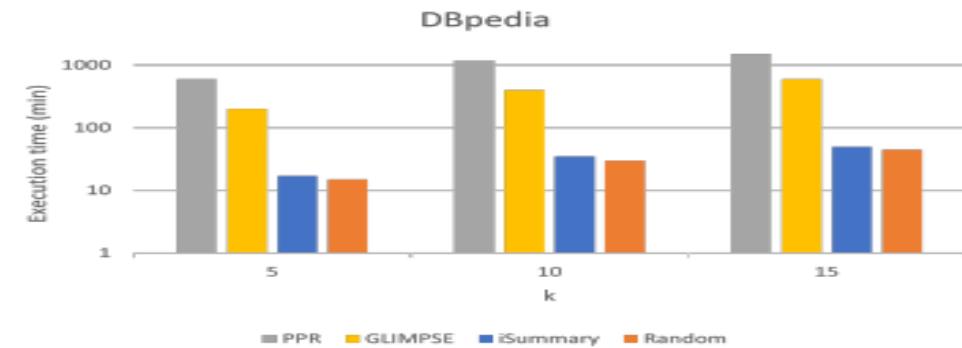


Fig. 6. Execution time for the various k and algorithms

5B: KGs and Inductive Learning

森田研究室

pyRDF2Vec: A Python Implementation and Extension of RDF2Vec

Bram Steenwinckel, Gilles Vandewiele, Terencio Agozzino, and Femke Ongenae/IDLab, Ghent University

- 一言でいうと
知識グラフから意味的にベクトル表現に変換するRDF2VecアルゴリズムのPython実装であるpyRDF2Vecを紹介している。
- Resource Type
Software
- 動機
RDF2VecのオリジナルのコードはJavaで書かれている。
- リソースの設計方針
FAIR原則に従って設計されており、人間と機械が読み取り可能な適切な説明が提供されている。
- 新規性
より大規模で複雑な知識グラフに対応可能な点である。
- リソースの再現性／活用実績
GitHub上で公開されており、ドキュメントも充実している。
- 品質
多数の拡張機能を備えており、幅広いタスクで適用可能である。
- リソースの可用性
URL: <https://github.com/IBCNServices/pyRDF2Vec>
- 次にすべきこと
今後も新しい拡張機能を追加することで、より多様なアプリケーションに対応できるようにすること。

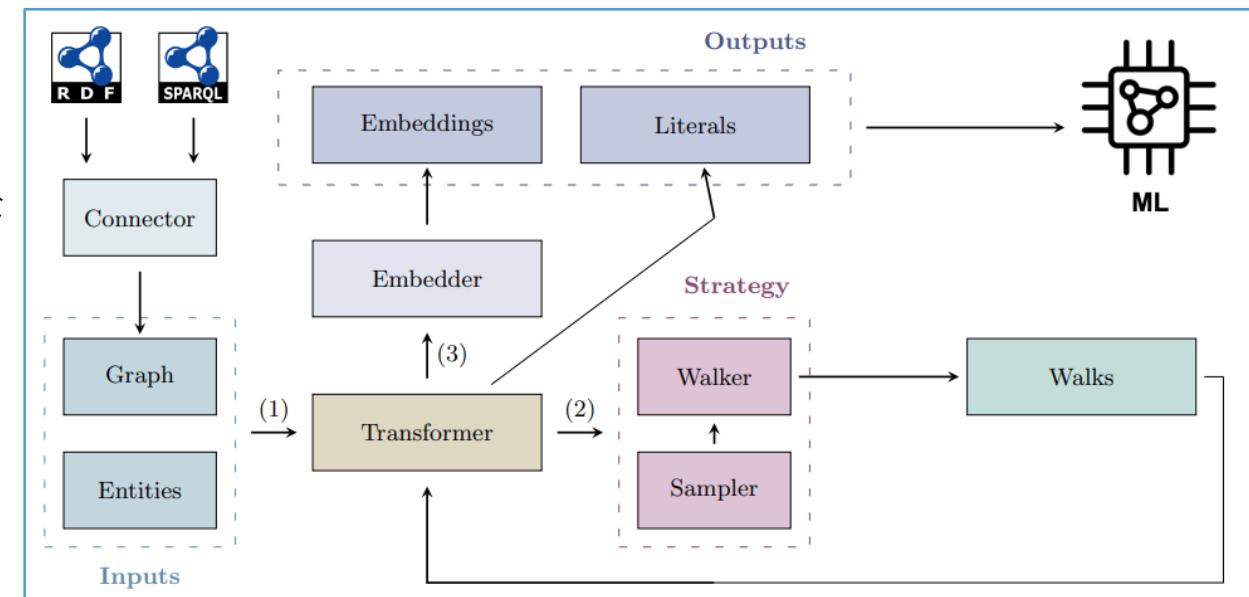
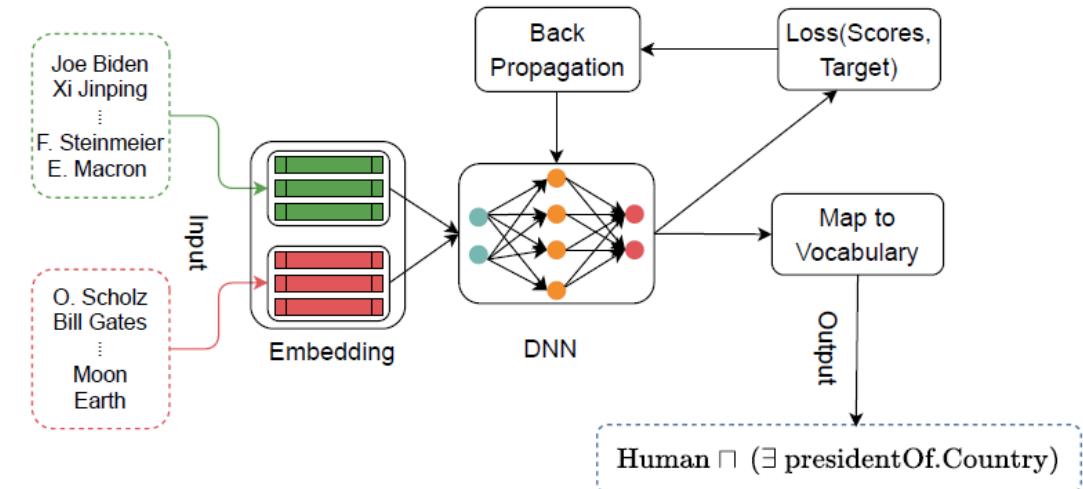


Fig. 2: Workflow of pyRDF2Vec.

Neural Class Expression Synthesis

N'Dah Jean Kouagou, Stefan Heindor, Caglar Demir, Axel-Cyrille Ngonga/Paderborn University, Germany

- 一言でいうと
クラス表現学習のための、ニューラルクラス表現合成器 (NCES) という新しいアプローチの提案
- 動機
スケーラビリティが欠如しており、適切なクラス表現を見つけるのに時間がかかる
- 手法
クラス表現学習を翻訳問題として捉え、機械翻訳のような方法で入力例からクラス表現を直接合成
- 背景
現状、クラス表現学習問題を解く方法最も優れているのは、ニーメント演算子や進化的アルゴリズムに基づく
- 結果
大規模な知識ベースにおいて全ての最先端の検索ベースのアプローチを凌駕することを示した
- 考察
学習時間を考慮すると本アプローチは、同じ知識ベースに対して多くの概念を学習するアプリケーションシナリオに適している
- 課題
知識ベースで獲得した知識を他の知識ベースに転送する手段の検討
本アプローチをALCHIQ(D)のような、より表現力の高い記述論理に拡張

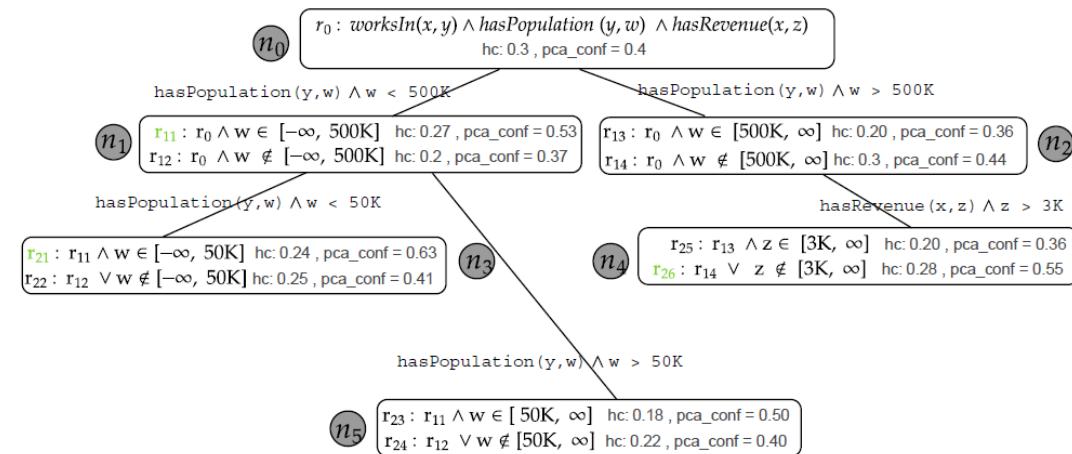


REGNUM: Generating Logical Rules with Numerical Predicates in Knowledge Graphs

Armita Khajeh Nassiri¹, Nathalie Pernelle^{1,2}, and Fatiha Saïs¹

/ ¹LISN, CNRS (UMR 9015), Paris Saclay University, Orsay F-91405, France, ²LIPN, CNRS (UMR 7030), University Sorbonne Paris Nord, France

- 一言でいうと
ルールマイニング・システムによってマイニングされたルールの肩の上に数値ルールを構築する新しいアプローチREGNUMの紹介
- 動機
金融、公衆衛生、ライフサイエンスなどの領域において有用な情報である数値ルールを考慮するため
- 手法
有意性を犠牲にすることなくルールの信頼性を高めるために、決定木を用いた数値の教師付き離散化を用いることで、KGからマイニングされたルール群を拡張
- 背景
多くのルールマイニングシステムは数値述語が多数の値を取り得るため、検索空間が膨大になることが課題
- 結果
拡張されたルールは平均品質が高く、知識グラフ補完タスクにおけるルールマイニングシステムの精度向上を支援
- 考察
決定木がより関連性の高い区間を提供することで、より多くの親ルールを充実させることができることを発見
- 課題
逐次的な被覆アプローチといった、制約を得るために別の方法の探求 クエリの実行時間を改善するためにインメモリデータベースの使用



Industry Track 1

kgyanagi

Wisdom of the Sellers: Mining Seller Data for eCommerce Knowledge Graph Generation

Petar Ristoski, Sathish Kandasamy, Aleksandr Matiushkin, Sneha Kamath, and Qunzhi Zhou/
eBay Inc., San Jose, USA

- 一言でいうと
eCommerce検索エンジンにおける、KGの導入
- 動機
eCommerce検索エンジンにおいて、ユーザの意図と製品在庫との間にミスマッチが生じる。
- 課題
semantic query understandingのための有用な知識ベースの生成
- 手法
カテゴリ別に分割された、売り手によって提供されるリスト中の共起アスペクト - 値ペアに基づく有向加重グラフを構築
この重みは、グローバルな人気によって正規化され、一緒に発生する頻度の高いアスペクトと値のペアにより高い関連性を与える
- 評価
数百万人のユーザを対象としたA/Bテストを実施
基本的なe-Commerceアプリケーションで活用 : Semantic Query Expansion, KG-enhanced Query Reformulation, Multi-Faceted Item Recommendation, Listing Autocomplete and Validation
- 考察
A/Bテストにより、検索放棄率が統計的に有意に低下し、低想起検索セッションが減少し、購入製品が有意に増加することを確認
- 知見
A/Bテストにより、検索放棄率が統計的に有意に低下し、低想起検索セッションが減少し、購入製品が有意に増加することを確認

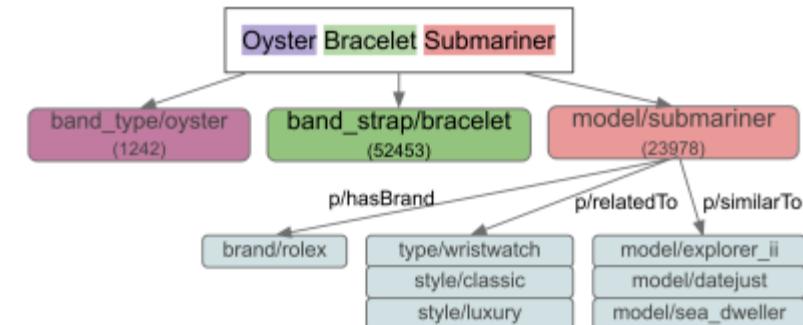


Fig. 1: KG-based semantic query understanding

Supplier Optimization at Bosch with Knowledge Graphs and Answer Set Programming

Cuong Xuan Chu, Mohamed H. Gad-Elrab, Trung-Kien Tran, Marvin Schiller, Evgeny Kharlamov, and Daria Stepanova/
Bosch Center for Artificial Intelligence, Renningen, Germany

- 一言でいうと

自動車産業における部品サプライヤーを知識グラフに基づいて最適化するために、アンサーセットプログラミング (ASP) を使用した経験のレポート

- 動機

自動車業界における、コストを低く抑えながら顧客の要求を満たすようにサプライヤを最適化するという課題

自動車業界におけるサプライヤ最適化の課題は、品質、信頼性、コスト、納期などのさまざまな基準に基づいて、自動車部品およびコンポーネントの最適なサプライヤを特定

- 課題

ナレッジグラフは、複雑なサプライチェーンをモデル化するための効果的なツールであることが証明されているが、最適化に使用できるツールは限られている。

KGのサイズが大きく、全データ上でASPソルバを直接呼び出すことは実行不可能。

- 手法

アンサーセットプログラミング (ASP) を使用
最適化に必要な関連事実のみを抽出

- 評価

実データを用いてサプライヤ最適化サービスを評価。筆者らのモデル化フレームワークの効率と有効性の両方を示した。

仕入先最適化サービスにより、仕入先を171社から3社に減らすことができた。

最も二酸化炭素排出量の多いサプライヤーの最適な数がわずか15秒で計算され、実際の産業環境での当社のサービスの有用性が実証された。

- 考察

製品に対する評価において、筆者らのモデル化フレームワークの効率と有効性の両方を示した。

- 知見

製品に対する評価において、筆者らのモデル化フレームワークの効率と有効性の両方を示した。

```

graph LR
    KG[Knowledge Graphs] --> KGQ[KG Query]
    subgraph KGQ [KG Query]
        - sub-parts
        - current suppliers
        - potential suppliers
        - supplier meta info
        - carbon footprint
    end
    KGQ --> ASPR[ASP Representation]
    subgraph ASPR [ASP Representation]
        - ASP grounding facts
        - ASP rules
        - ASP constraints
        - ASP objectives
    end
    ASPR --> ASPSR[ASP Solver Running]
    subgraph ASPSR [ASP Solver Running]
        - ASP library
        - ASP parameters
    end
    ASPSR --> OSG[Optimized Supplier Graph]
    OSG

```

Fig. 1. Architecture overview of the supply chain optimization service.

Ontologies for Formalizing the Process of Configuring and Deploying Building Management Systems

Hervé Pruvost and Andreas Wilde/

Fraunhofer Institute for Integrated Circuits (IIS), Division Engineering of Adaptive Systems (EAS), Münchener Straße 16, 01187 Dresden, Germany

- 一言でいうと

ビル管理システムの設計と構成のプロセスを自動化するために知識グラフを活用する方法を提案

- 動機

建物は世界のエネルギー消費量の約1/3を占め、この30%は、暖房、換気、空調システム (HVAC) の誤動作などによる浪費

ビル管理システム (BMS) は、ビルのエネルギーシステムの最適制御と連続監視を支援するが、その設計と構成のプロセスは複雑。

- 課題

この分野の既存のオントロジー (ifcOWL、BOT、QUDTオントロジー) では建物の操作段階における記述は不可能

- 手法

操作段階における建物の完全な記述を可能とするオントロジーの開発。

Senseオントロジーは、専門知識を形式化するために論理的な公理と規則を符号化。故障検出のためのBMS機能のセットアップのために、SWRLでルールを記述。

- 評価

メタデータ処理と、それに続きデータ処理からなる2段階のワークフローを実現するプロトタイプを開発。

- 考察

BMS機能が、エネルギーシステムのオントロジー記述から導出できることを証明。

- 知見

BMS機能が、エネルギーシステムのオントロジー記述から導出できることを証明。

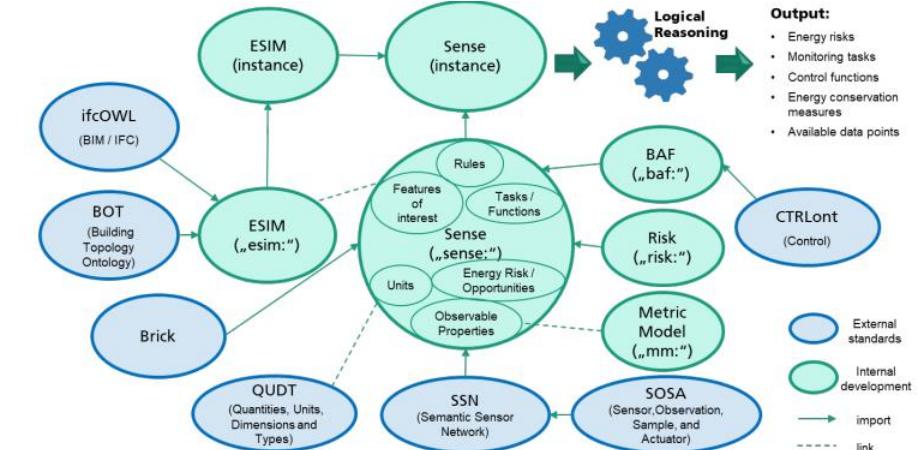


Fig. 1. Ontology system

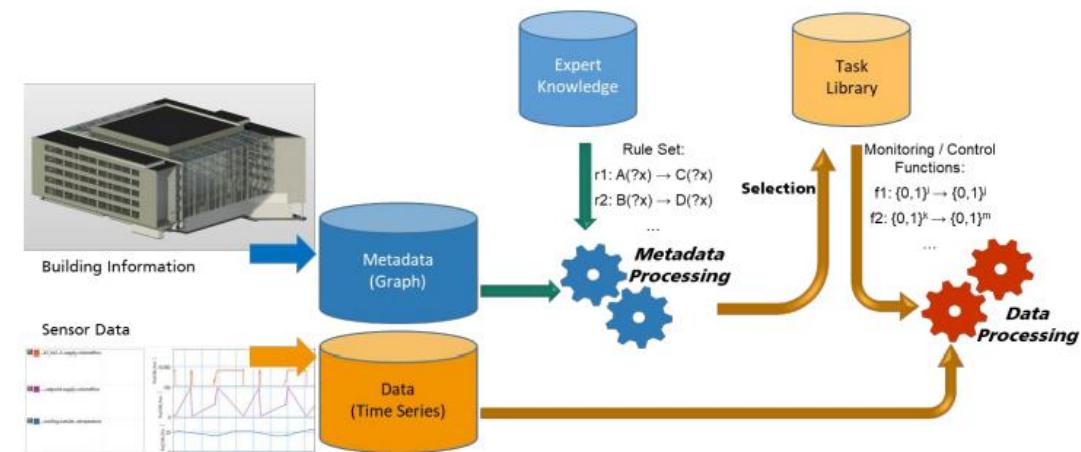


Fig. 3. Workflow for automated building energy system monitoring and control

Towards the Deployment of Knowledge Based Systems in Safety-Critical Systems

De-Grancey Florence and Audouy Amandine/
Thales AVS France, SAS, France

- 一言でいうと
コックピットシステム内のクルーアシスタントを作成するための知識ベースシステムの活用
- 動機
safety-critical system (SCS) におけるAI技術の導入、特に機械学習が広く研究されている。知識ベースシステム (KBS) は、既存のドメイン知識に価値を付加したり、因果的な説明を提供したりする能力があり魅力的。
- 課題
 1. タスクに適した知識ベースの設計の容易化
 2. ブラックボックスの検証と検証の強化
 3. 許容可能で制限された実行時間
- 手法
 1. 航空学ドメインオントロジーSESAR BEST 2 AIRMから出発して、期待されるタスクに必要な要素だけを選択するようになぞらえを改良。OntoDebug 3を用いてエラーチェックを行い、OntoMetrics 4を用いてメトリックを監視。
 2. ブラックボックスのテスト方法を探求し、テストが手動で作成されたABoxや、それらのABoxに関する一連のバリエーションによって定義されるいくつかのテストベースのキャンペーンを設計。
 3. 各テストを少なくとも100回起動して実行時間の統計を測定。
- 評価
推論タスクの実行時間の統計を測定。(3)
- 考察
query task以外のタスクの実行時間には大きなばらつきがあった
- 知見
安全臨界システムとしてのKBSの開発は、大量の検証活動や限定推論時間のような新しいニーズをもたらした

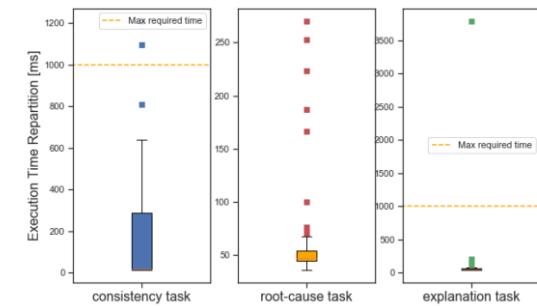


Fig. 1. Repartition of measured execution time and required execution time (1000ms)

Industry Track 2

Nagano

Addressing the Scalability Bottleneck of Semantic Technologies at Bosch

Diego Rincon-Yanez^{1,2}, Mohamed H. Gad-Elrab¹, Daria Stepanova¹, Kien Trung Tran¹, Cuong Chu Xuan¹, Baifan Zhou^{3,4}, and Evgeny Karlamov^{1,3}

1: Bosch Center for AI, Germany
2: University of Salerno, Italy

3: SIRIUS Centre, University of Oslo, Norway
4: Oslo Metropolitan University, Norway

- **一言でいうと**

人間にもAIにも解釈可能な共通言語としてオントロジー／知識グラフを利用し、データ処理のパイプライン構築を自動化

- **動機**

製造業のスマート化に向けたデータ活用による意思決定の迅速化

- **課題**

多種多様なデータの統合、および迅速な計算処理（機械学習の活用？）が必要となり、スケーラビリティのボトルネックになっている

- **手法**

セマンティック技術（オントロジー、知識グラフ）と機械学習技術を融合したデータ処理パイプラインを構築
具体的な内容は文献[10,11]で報告

- **評価**

記載なし

- **考察**

記載なし

- **知見**

セマンティック技術は共通データスキーマの提供に加えて、異なるステークホルダー間のコミュニケーションに有用な共通言語を提供し、それがソリューション開発の効率化に大きく寄与している

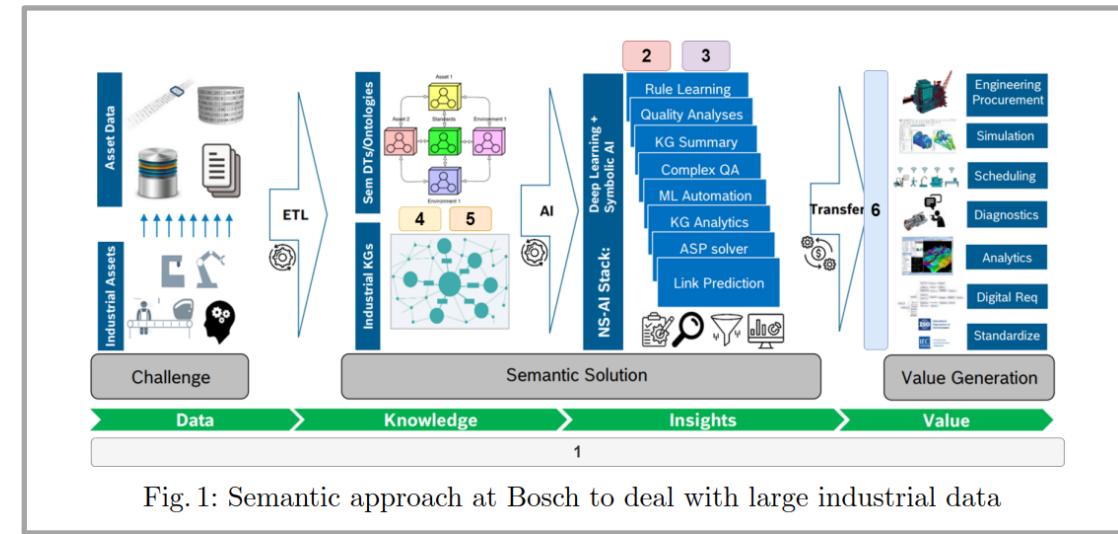


Fig. 1: Semantic approach at Bosch to deal with large industrial data

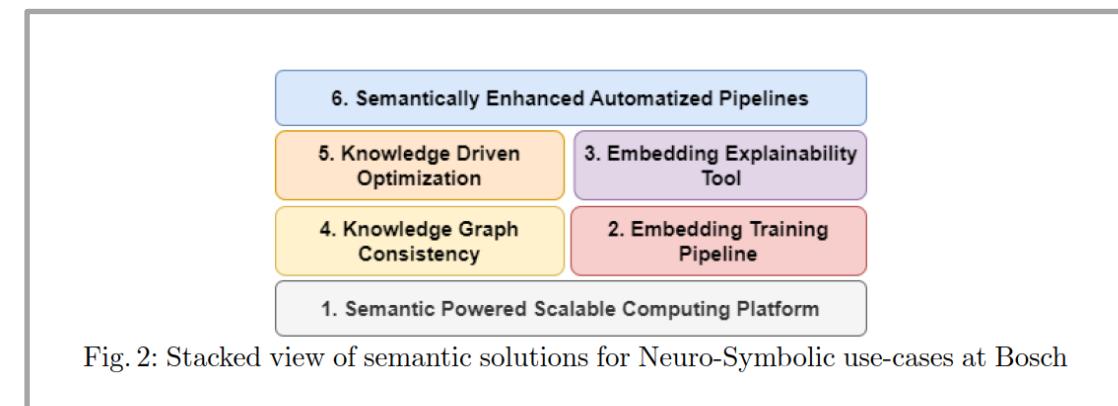


Fig. 2: Stacked view of semantic solutions for Neuro-Symbolic use-cases at Bosch

図1,2中の番号1から6はそれぞれ対応

Knowledge Injection to Counter Large Language Model (LLM) Hallucination

Ariana Martino, Michael Iannelli, Coleen Truong / Yext, USA

- 一言でいうと
○○○
- 動機
LLMに出力には誤りが含まれる可能性があり、企業によるLLM導入の妨げの1つになっている
- 課題
同上
- 手法
誤り情報の出力を軽減するため、知識注入（KI）と呼ばれる技術を利用する
- 評価
定性的なレビューでは、KIを用いて微調整された bloom-560mは、OpenAIの微調整されていない text-davinci-003モデルを上回った
- 考察
記載なし
- 知見
記載なし

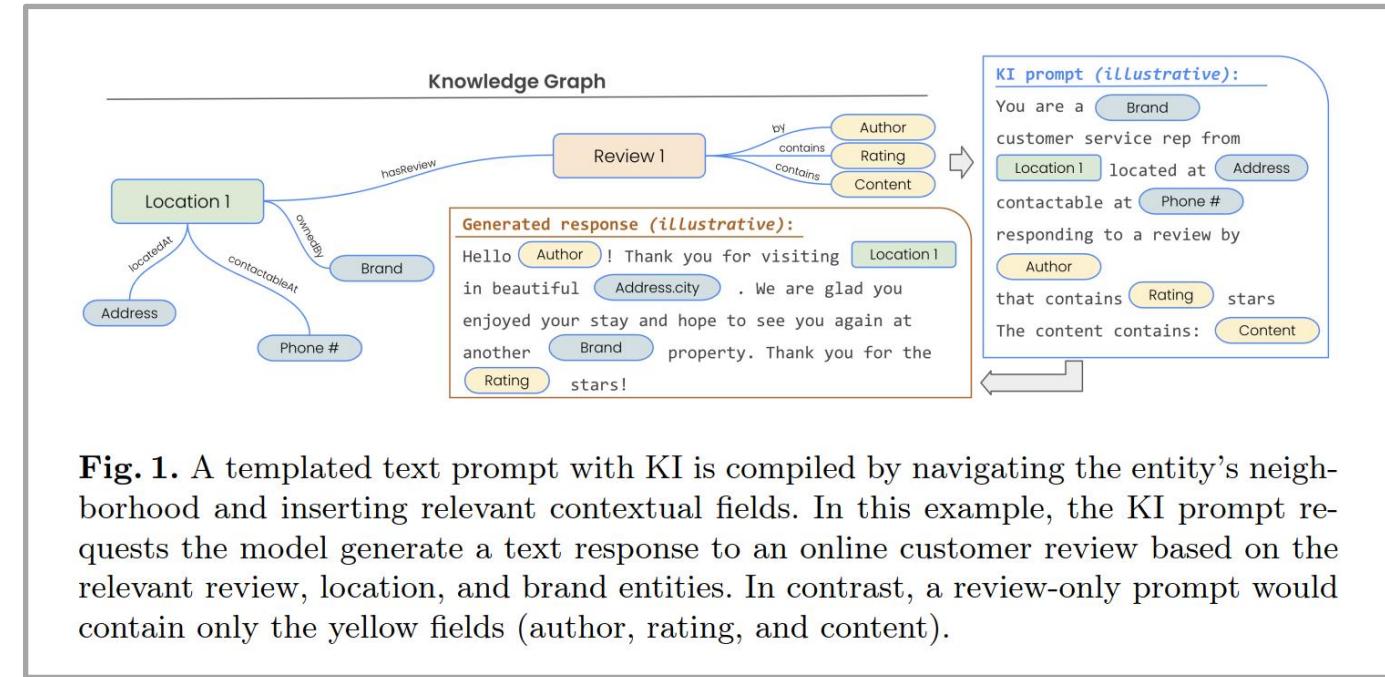


Fig. 1. A templated text prompt with KI is compiled by navigating the entity's neighborhood and inserting relevant contextual fields. In this example, the KI prompt requests the model generate a text response to an online customer review based on the relevant review, location, and brand entities. In contrast, a review-only prompt would contain only the yellow fields (author, rating, and content).

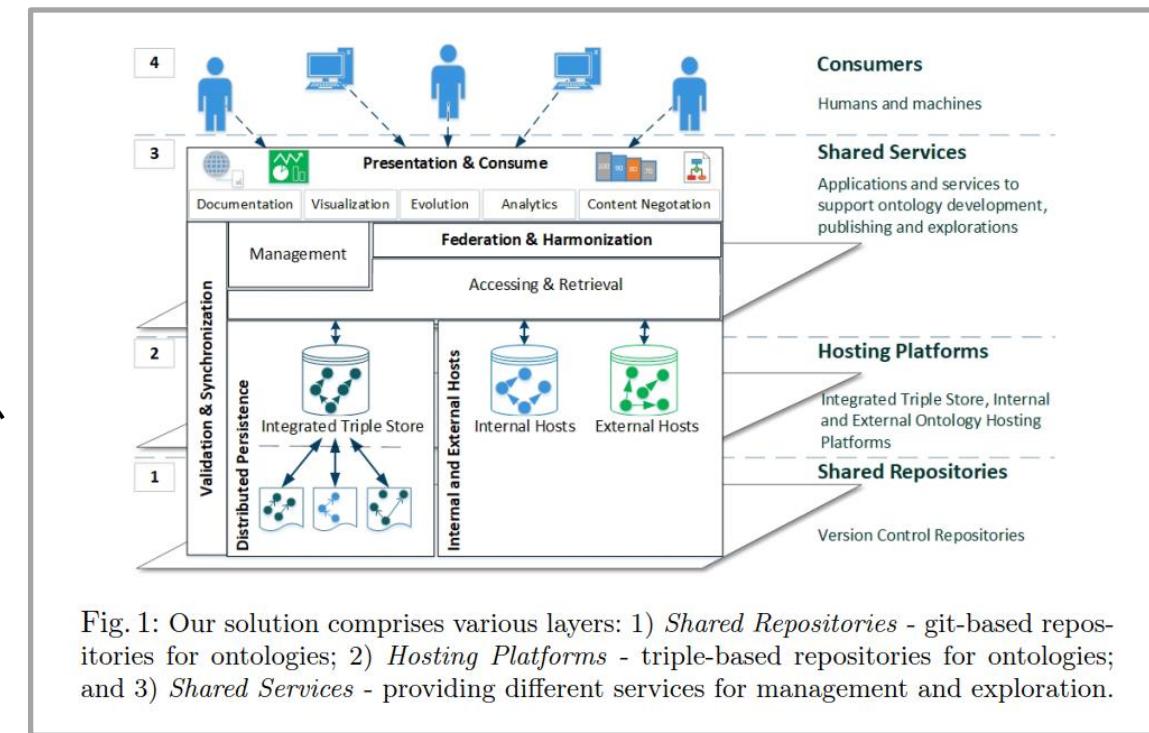
知識注入（KI）は、テキスト生成タスクに関するエンティティに関するコンテキストデータを、知識グラフからテキスト空間にマッピングしてプロンプトに含める。

A Source-agnostic Platform for Finding and Exploring Ontologies at Bosch

Lavdim Halilaj¹, Stefan Schmid¹, Khushboo Goutham Chand², Santhosh Kumar Arumugam², and Sahu Sajita Kumari²

1: Bosch Center for Artificial Intelligence, Germany
2: Bosch Global Software Technologies Private Limited, India

- 一言でいうと
オントロジーの再利用性を高めるプラットフォームの提案
- 動機
各産業分野において多様なデータの理解／統合／相互接続を目的に構築してきたオントロジーの再利用性を高めたい
- 課題
各プロジェクトで開発されたオントロジーは異なるプラットフォームにホストされることが多く、再利用が容易ではない
- 手法
オントロジー／トリプルストアの統一的アクセスと拡張性を両立可能にする、3層アーキテクチャで構成のプラットフォームを提案
- 評価
具体的な実装と評価の記載はないが、Bosch社内で既にプラットフォームが運用されており、500以上のオントロジーを提供し、100人以上のユーザが利用していると報告されている
- 考察
オントロジー開発の効率化にはドメイン専門家の理解が不可欠で、そのための多様なビューをプラットフォームで提供していく
- 知見
記載なし



ご発表ありがとうございました！