

第4回ナレッジグラフ勉強会

ISWC2023論文読み会

2023/12/20 17:00 – 19:00

主催：ナレッジグラフ若手の会

会場：オンライン

タイムテーブル

時間	コンテンツ	セッション	登壇者
16:45-	開場		
17:00-17:10	オープニング		事務局
17:10-17:13	発表1	Industry: Internet of Things (IoT) and Data Enrichment	Nagano
17:13-17:16	発表2	Industry: Data Management and Analysis	Nagano
17:16-17:19	発表3	7B: RDF Dataset Management	koji6_fujiwara
17:19-17:21	発表4	1B: Search, Retrieval and SPARQL I	koji6_fujiwara
17:21-17:24	発表5	5A: Search, Retrieval and SPARQL II	山口研究室
17:24-17:27	発表6	Industry: Semantic Data and Metadata	山口研究室
17:27-17:30	発表7	2B: Knowledge Extraction	whirota
17:30-17:33	発表8	9B: Entity Alignment	whirota
17:33-17:36	発表9	2A: Knowledge Engineering with Large Language Models	古崎研究室
17:36-17:39	発表10	10A: Ontology engineering and ontology patterns	古崎研究室
17:39-17:41	発表11	1A: Ontologies and Knowledge Graphs I	森田研究室
17:41-17:44	発表12	7A: Ontologies and Knowledge Graphs III	森田研究室
17:44-17:48	発表13	6B: Ontologies and Knowledge Graphs II	nomotom
17:48-17:51	発表14	8B: Ontologies and Knowledge Graphs IV	marshma84093472
17:51-17:54	発表15	3B: Internet of Things	marshma84093472
17:54-17:57	発表16	Industry: Ontologies and Knowledge Graphs II	Yasunori
17:57-18:00	発表17	10B: Linked Data in action	Yasunori
18:00-18:03	発表18	4A: Link Prediction I	ichise
18:03-18:06	発表19	6C: Link Prediction II	ichise
18:06-18:09	発表20	3A: Knowledge Graph Embeddings I	N.Yoshimaru
18:09-18:12	発表21	9A: Knowledge Graph Embeddings IV	N.Yoshimaru
18:12-18:15	発表22	6A: Knowledge Graph Embeddings II	kgyanagi
18:15-18:18	発表23	8A: Knowledge Graph Embeddings III	Shusaku Egami
18:18-18:21	発表24	7C: Temporal Reasoning	Shusaku Egami
18:18-18:25	クロージング		事務局
18:25-19:00	交流会（自由参加）		

Industry: Internet of Things (IoT) and Data Enrichment

Nagano

Automated verification of measurement precision for Internet-of-Things equipment

Mikkel Haggren Brynildsen^{1,†}, Maja Milićić Brandt^{2,†}, Johan Wilhelm Klüwer^{3,†},
Caitlin Woods^{4,†} and Melinda R Hodkiewicz^{4,*†}

¹Grundfos Data & AI, Poul Due Jensens Vej 7, 8850 Bjerringbro, Denmark

²Siemens AG, Technology, Otto-Hahn Ring 6, 81739 Munich, Germany

³DNV, Veritasveien 1, 1363 Hovik, Norway

⁴School of Engineering, University of Western Australia M050, Perth, 6009 Australia

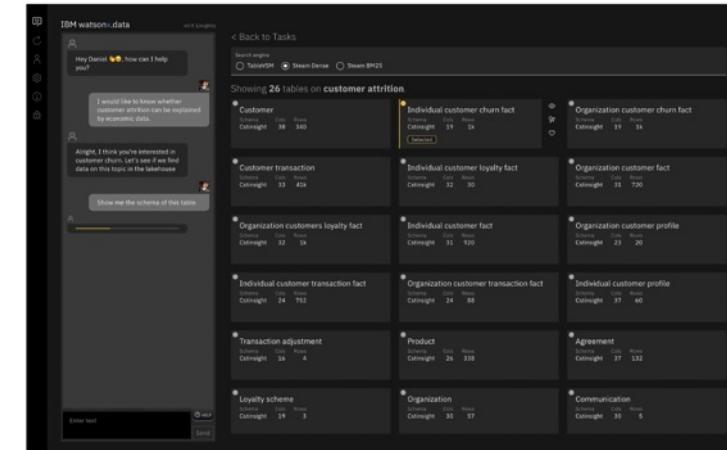
- 一言でいうと
産業上位オントロジーIDOを利用してIoT機器の計測精度
IDO: Industrial Data Ontology
- 動機
IoT機器の製品バリューチェーン（OEM）では、サプライヤーから提供されるセンサ監視ソフトウェアの更新が顧客へ提供される機器の計測精度に悪影響を及ぼさないことを確認するための保証プロセスを確立する
- 課題
一般にはデータシート等を使用して関係者間で技術情報が交換されるため、データ相互運用性の確認には多大な人手作業が必要
- 手法
 - ポンプ本体（ハード、ソフトの両方）、及びポンプで送られる流体を対象に、センサーの分解能、及び計測されたデータの精度に関する情報と紐づけ、IDOに基づいてセマンティックモデルを構築
 - IDOはISO TR 15926-14の進化版として、ISO TC 184/SC4 産業データ委員会で標準規格化中
- 評価
ACME社のポンプを題材に、メーカー2社を交えて、ファームウェアの更新情報と紐づけてデータ交換できることを確認
- 考察
ソリューションの拡張性、トレーサビリティの確保、品質管理における人的依存の排除
- 知見
記載なし

Conversational GUI for Semantic Automation Layer

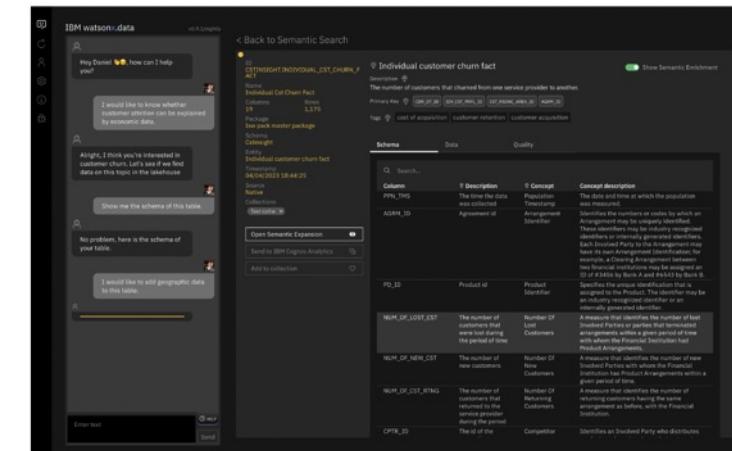
Daniel Karl I. Weidle*, Gaetano Rossiello, Gregory Bramble, Abel N. Valente, Sugato Bagchi, Faisal Chowdhury, Mauro Martino, Nandana Mihindukulasooriya, Haritha Ananthakrishnan, Hendrik Strobelt, Hima Patel, Alfio Gliozzo, Owen Cornea, Ankush Gupta, Sameep Mehta, Manish Kesarwani, Horst Samulowitz, Oktie Hassanzadeh, Arvind Agarwal and Lisa Amini

IBM Research

- 一言でいうと
業務／業界で利用される表データを意味づけして取り込むための対話型GUIを提案
- 動機
データ管理やデータサイエンスにおいて用語集、用語マッピングが不可欠
- 課題
既存のセマンティック層（セマンティック技術と解釈）は多大な人手作業を必要とする
- 手法
 - ・大規模言語モデル（watson.ai）を利用して、表データに意味づけを行う
 - ・構成要素：1)セマンティック検索、2)セマンティックエンリッチメント（表への意味付け）、3)用語集の統合支援、4)データ品質プロファイリング
- 評価
ユースケース：ビジネスアナリストが銀行顧客の離反が経済データで説明できるかどうかを調査したいと考えている。チャットで依頼すると、エージェントが関連するテーブルデータを検索。次に、テーブルのスキーマ調査を指示すると、主キーを推定し、各カラムにタグ付けを実施。さらに、ローカルに持つ失業率のCSVファイルと意味的に統合し、結合テーブルの品質レポート（重複や外れ値の有無）を生成。
- 考察
記載なし
- 知見
記載なし



セマンティック検索の出力例。意味的関連性にもとづいて表データがランキングされている



表データに対するセマンティックエンリッチメントの出力例。各カラムの説明が生成され、タグが付与されている

Building an Industrial Ontology Engineering Platform

Mikkel Haggren Brynildsen^{1,*†}, Claus Jakobsen^{2,†}, Nicolaj Abildgaard^{2,†} and Caitlin Woods^{1,*†}

¹Grundfos, Poul Due Jensens Vej 7, 8850 Bjerringbro, Denmark

²Delegate, Hummeltofevej 49, 2830 Virum, Denmark

- 一言でいうと
産業オントロジーのエンジニアリングを効率化するツールをIOEP開発
- 動機
産業界のデータプロジェクトはデータの量と複雑さが増大しており、その制御のために共通データモデルが必要であり、オントロジーは有用な仕組みの一つ
- 課題
オントロジーエンジニアリングの組織的課題（スケーラビリティに課題と解釈）
- 手法
下記の目標を満たすオントロジーエンジニアリングのツールを構築
G1) ドメイン専門家を新しいオントロジーモデルの所有者、貢献者とする
G2) オントロジーエンジニアの作業負荷を軽減する
G3) エンタープライズデータガバナンス要件を満たす
G4) データの新規作成と既存ソースからの取り込みの両方をサポートする

評価

現在、異なる3件のプロジェクト向けに3つのオントロジーを開発中。

考察

オントロジー設計パターンを利用することで、オントロジーの複雑さを抽象化でき、1) オントロジー設計の再利用性が向上、2) ドメイン専門家による設計が容易化

知見

- ・表形式であればドメイン専門家でもクラス階層のモデル化は容易。
- ・今後の課題：1)複雑なパターンを表形式で表現、2)オントロジーガバナンスのオープンスタンダード開発

IOEPツールのスナップショット

ワークスペースの
スナップショット
検索

The screenshot shows the IOEP tool's interface. At the top, there's a navigation bar with 'GRUNDFOS' logo, 'Ontologies / Workspaces / demo'. On the right, it says 'Caitlin Woods'. Below the navigation, there's a 'Workspace' dropdown set to 'demo' (marked with a circled 5). A search bar labeled 'Search' (marked with a circled 4) is followed by a list of source tables: 'GPCunits', 'lis14-class-data-table', 'lis14-object-data-table', 'ontology-metadata', and 'RDL-annop-data-table'. At the bottom left are 'PUBLISH' and 'BUILD' buttons (marked with a circled 3). The main area is titled 'RDLQuantityDatum' and contains a table with columns: 'quantityDatum', 'label', 'superclass', 'modifiedBy', 'modifiedDate', 'approvalStatus', 'modelStatus', and 'change'. There are two rows of data (marked with circled 1 and 2). Row 1: 'quantityDatum' is 'gf:a8b824d2-6fcf-43e2-a3bc 0063d16f57d3', 'label' is 'temperatureDatum', 'superclass' is 'lis:ScalarQuantityDatum', 'modifiedBy' is 'Mikkel Haggren Brynildsen', 'modifiedDate' is '2023-05-15T12:24:07.6552539Z', 'approvalStatus' is 'Define', 'modelStatus' is 'Development', and 'change' is 'MIHB t'. Row 2: 'quantityDatum' is 'gf:020312df-906a-4eec-9603-dr121078cher', 'label' is 'frequencyDatum', 'superclass' is 'lis:ScalarQuantityDatum', 'modifiedBy' is 'Mikkel Haggren Brynildsen', 'modifiedDate' is '2023-05-17T06:17:18.8176366Z', 'approvalStatus' is 'Define', 'modelStatus' is 'Development', and 'change' is 'MIHB t'. The table has a 'Columns (8/13)' dropdown and a 'Filter table' button at the top right.

quantityDatum	label	superclass	modifiedBy	modifiedDate	approvalStatus	modelStatus	change
gf:a8b824d2-6fcf-43e2-a3bc 0063d16f57d3	temperatureDatum	lis:ScalarQuantityDatum	Mikkel Haggren Brynildsen	2023-05-15T12:24:07.6552539Z	Define	Development	MIHB t
gf:020312df-906a-4eec-9603-dr121078cher	frequencyDatum	lis:ScalarQuantityDatum	Mikkel Haggren Brynildsen	2023-05-17T06:17:18.8176366Z	Define	Development	MIHB t

オントロジーの
構築と公開

識別子(IRI)の自動生成

ガバナンスに関する属性の自動生成

Industry: Data Management and Analysis

Nagano

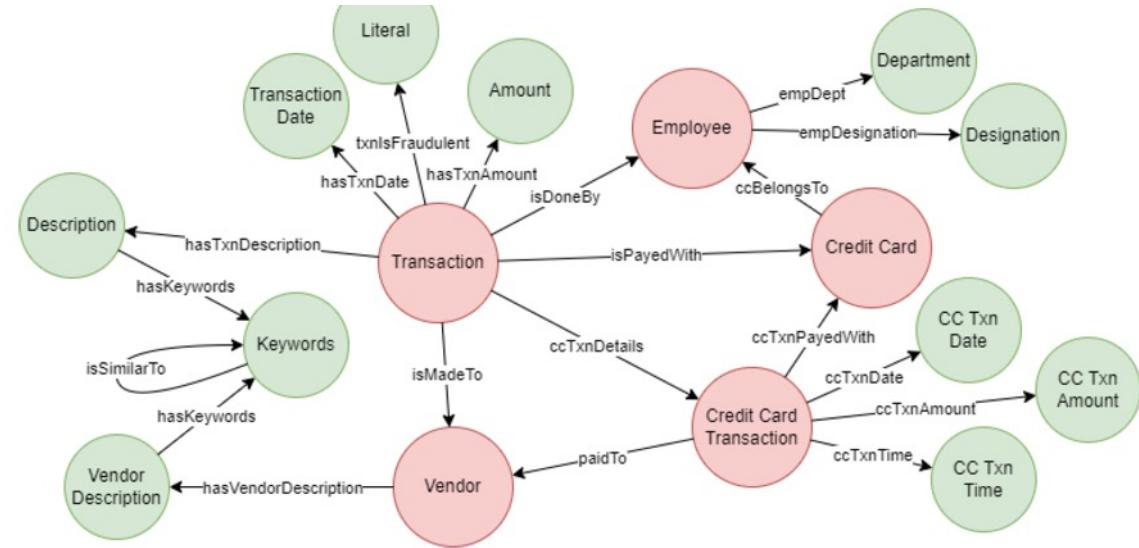
Graph Representational Learning for Internal Audit

Pai, Sumit^{1,*†}, Singh, Vivek Kumar^{1,†}, Gupta, Sanvi¹, Chavali, Pavani¹,
Siddhartha, Siddhartha¹, Bowen, Edward² and Tiyagura, Sunil Reddy¹

¹Deloitte & Touche Assurance & Enterprise Risk Services India Private Limited

²Deloitte & Touche LLP

- 一言でいうと
内部監査において、ナレッジグラフと表現学習を利用して経費処理の不正を特定
- 動機
組織ガバナンスにおける内部監査員の業務効率化
- 課題
 - ルールベース方式は事前定義ルールの範囲でしか特定できずスケールしない
 - 標準的な外れ値検出手法（IF、AEなど）は、分布変化やラベル不均衡のあるデータに対して機能しない、データ間の関係性を活用できない
- 手法
 - 表形式の経費記録データをナレッジグラフ（右図）へ変換。
テキスト列のキーワードはBERT単語埋込を利用して類似するどうしを接続
 - 不正トランザクションの分類タスクとして定式化、ナレッジグラフ埋込（KGE）で学習
- 評価
 - 少数のトランザクション(<1%)に不正ラベルを付与したデータセットを作成
 - KEGが外れ値検出手法（IF、AE）より精度が上回ることを確認
- 考察
データ間の関係性をモデリングしたことが精度向上に寄与した（詳細は不明）
- 知見
未知のシンボリックノード（不明）が存在するため、新規データに対して再学習が必要なる点が課題



ナレッジグラフの例。赤はプライマリーノード、緑は属性

Model	Precision	Recall	F1
IF	0.25	0.45	0.32
AE	0.32	0.55	0.40
KGE	0.59	0.55	0.57

精度評価の結果。

IE: Isolation Forests, AE: Auto Encoders

Railway track video Knowledge Base

Kai Herbst^{1,*}, Rene Krieg¹ and Dirk Friedenberger¹

¹DB Systel GmbH, Frankfurt, Germany

- **一言でいうと**
鉄道車両から撮影した線路の動画データを、鉄道ネットワークのデータ、車両運行データと紐づけて、ナレッジベースを構築
- **動機**
鉄道ネットワーク保守業務での動画データ活用の容易化、運転士の教育
- **課題**
鉄道車両から撮影した線路等の動画を記録、処理、提供しているが、動画内の特定のポイントへジャンプできなかった
例：ある駅の入口信号から出口信号までを通過した区間の動画
- **手法**
 - 鉄道ネットワーク上の駅、信号などの要素から構成された独自のオントロジーを構築
 - 鉄道車両運行監視制御システムLeiDisから取得した車両のリアルタイム情報（日時、GPS）とリンクさせてナレッジグラフを生成
- **評価**
動画内の要素（駅や信号など）に対してSPARQLで検索が可能になった（PoCレベルと推測される）
- **考察**
参考にしたオントロジー構築方法論：Ontology Development 101[1], Generic Ontology Design Patterns[2]
- **知見**
今回、各要素の位置と車両の通過時刻は時刻表を参照。動画からの自動認識が今後の課題



ナレッジグラフに対してSPARQLクエリを実行し、検出した信号の例

Unleashing the Potential of Data Lakes with Semantic Enrichment Using Foundation Models

Nandana Mihindukulasooriya[†], Sarthak Dash, Sugato Bagchi, Faisal Chowdhury, Alfio Gliozzo, Ariel Farkash, Michael Glass, Igor Gokhman, Oktie Hassanzadeh, Nhan Pham, Gaetano Rossiello, Boris Rozenberg, Yehoshua Sagron, Dharmashankar Subramanian, Toshihiro Takahashi, Takaaki Tateishi and Long Vu

IBM Research

- 一言でいうと
表データのテーブル名と列名だけからメタデータを自動生成し、列から概念へのマッピングを自動付与するプロセスを提案
- 動機
多くの組織が異種データを含むデータレイクを管理している。データレイク中の表データに対して適切なメタデータが不足すると、データレイクがデータ沼に変化し、重要な組織タスクに関連するデータを見つけることが困難になる
- 課題
表データへのセマンティックエンリッチメントに関する学術研究は、オープンなナレッジグラフ（DBpediaやWikidataなど）へのリンク付与を対象としている一方、企業では以下の課題があり、多くの場合は表データのメタデータ（テーブル名と列名）のみを利用してメタデータを強化する必要がある
 - テーブル名、列名にはデータ所有者を表すコードや頭文字などの短縮文字列などが使われる
 - テーブル名や列名へのアクセスのみを許可し、実際のデータ（セル値）は許可していない
 - 組織独自のエンティティが含まれるため、オープンなナレッジグラフへリンクを付与できない
- 手法
 - 入力：表データのメタデータ（テーブル名、列名）、組織にとって関心のある概念を定義するビジネス用語集
 - プロセス
 - (a) 列名を拡張：短縮文字列から意味のある列名を生成する
 - (b) 表データのメタデータを強化：オープンデータを利用してメタデータ生成モデルを学習する
 - (c) 各列からビジネス用語集の概念へのマッピング（列タイプアノテーション）：列のメタデータ表現とビジネス用語集の用語表現の類似性を計算
- 評価／考察／知見
記載なし

ISWC2023サーベイ

7B: RDF Dataset Management

1B: Search, Retrieval and SPARQL I

藤原浩司(東芝)

7B: RDF Dataset Management

Dense Re-Ranking with Weak Supervision for RDF Dataset Search

Qiaosheng Chen, Zixian Huang, Zhiyang Zhang, Weiqing Luo, Tengteng Lin, Qing Shi and Gong Cheng

- 一言でいうと
RDFデータセット検索において高密度再ランキングを適用する
- 動機
既存のRDFデータセット検索は、スパースモデルを採用している
- 手法
遠隔監視に基づく方法と自己訓練に基づく方法の2種類の手法で
モデルを弱く監視する
- 背景
既存の高密度再ランキングは主に文書検索のためのものであり、RDFデータセット検索には適用されていない
- 結果
従来のスパース検索と比較して、結果が最大17%改善された
- 考察
RDFデータセット検索のタスクを、事前に訓練された言語モデルと高密度テキスト検索における最新の研究と結びつける
- 課題
RDFデータのサンプリング手法の別の手法による評価
コントラスト学習の適用
RDFデータセットだけでなく、より一般的なデータセット検索へアプローチを拡張

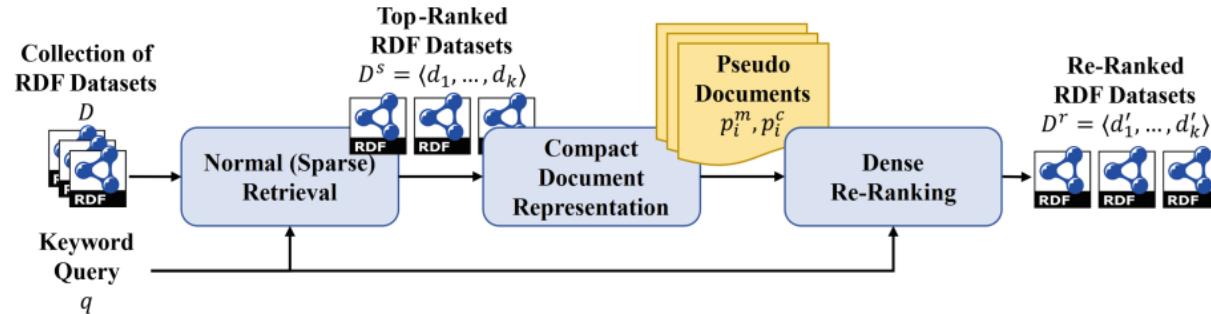


Fig. 1. Our retrieval-then-reranking approach for RDF dataset search.

Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology

Clement Jonquet, John Graybeal, Syphax Bouazzouni, Michael Dorf, Nicola Fiore, Xeni Kechagioglou, Timoth Redmond, Ilaria Rosati, Alex Skrenchuk, Jennifer L. Vendetti, Mark Musen and Members Of The Ontoportal Alliance

- 一言でいうと
オントロジーの管理を行うサービスOntoPortalの提案
- Resource Type
software tools/services
- 動機
オントロジーをFAIR(Findable, Accessible, Interoperable, and Reusable)にする
- リソースの設計方針
特定の分野に依存せず、多くのオントロジーを管理・利用可能にする
- 新規性
汎用オントロジリポジトリとして必要な各種機能を提供する
- リソースの再現性／活用実績
過去3年で135件の利用登録
- 品質
FAIRを推進している
- リソースの可用性
GitHub : <https://github.com/ontoportal>
デモ : [https://demo.ontoportal.org./](https://demo.ontoportal.org/)
- 次にすべきこと
技術的な機能改善(多言語化、Docker化など) / 検索、注釈、推薦などのサービス結果のフェデレーションに対する調整

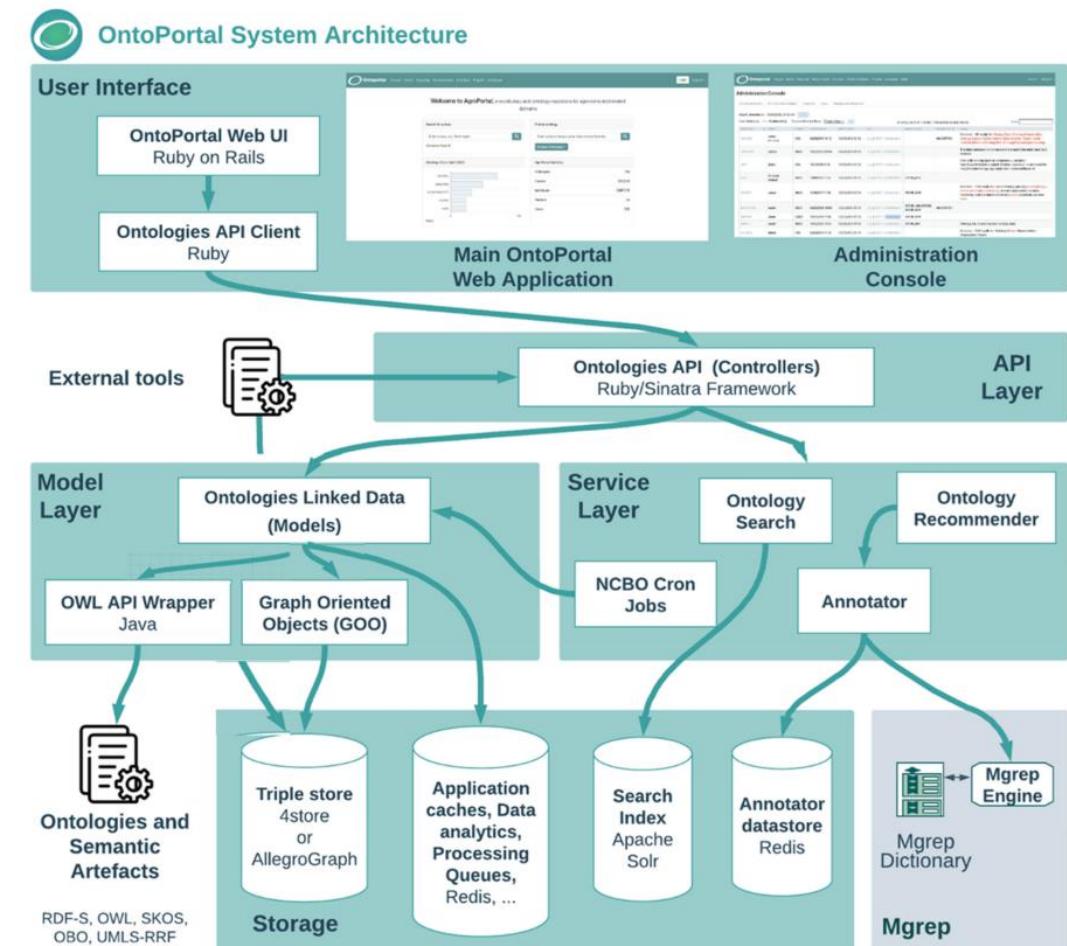


Fig. 2. OntoPortal system architecture.

VOYAGE: A Large Collection of Vocabulary Usage in Open RDF Datasets

Qing Shi, Junrui Wang, Jeff Z. Pan and Gong Cheng

- **一言でいうと**

RDFの語彙の使用状況分析のためのコレクション、VOYAGEの提案

- **Resource Type**

datasets

- **動機**

オープンRDFデータに対する一般的な語彙の分析は行われていない。

- **リソースの設計方針**

68,312のデータセットから62,864のクラスと842,745のプロパティを抽出し、50,976個に分類した。
エンティティ記述における共起パターンをの767,976個抽出した。

- **新規性**

語彙や用語の使用方法や各RDFデータセットから抽出された共起パターンを提供する。

- **リソースの再現性／活用実績**

紹介されている分析以外にも様々なシナリオに活用できる

- **品質**

各JSONファイルの構造を説明が公開されている

- **リソースの可用性**

データ : <https://zenodo.org/record/7902675>

コード : <https://github.com/nju-websoft/VOYAGE>

- **次にすべきこと**

1年、もしくはそれ以上の頻度で更新を続ける

Table 1. Statistics about Data Collection (Notes: ODP catalogues may overlap. Inaccessible ODPs/datasets and non-RDF datasets are not counted.)

ODP catalogue	#ODP	(%)	#dataset	(%)	#triple	(%)
CKAN	109	(18.89%)	15,858	(22.00%)	397,404,207	(40.96%)
DataPortals.org	110	(19.06%)	25,341	(35.15%)	555,050,054	(57.21%)
DKAN	36	(6.24%)	3,007	(4.17%)	14,345,698	(1.48%)
Open Data Portal Watch	135	(37.40%)	37,407	(51.89%)	689,106,507	(71.02%)
Socrata	398	(68.98%)	55,653	(77.20%)	427,739,164	(44.09%)
LOD Cloud	1	(0.17%)	308	(0.43%)	128,902,453	(13.29%)
Total	577	(100.00%)	72,088	(100.00%)	970,258,378	(100.00%)
After deduplication			68,312		920,501,102	

1B: Search, Retrieval and SPARQL I

Optimizing SPARQL Queries with SHACL

Ratan Bahadur Thapa and Martin Giese

- **一言でいうと**
SHACLによりRDFグラフに制約が適用されている場合に、そのRDFグラフに対して実行されるクエリを最適に書き換える。
- **動機**
解決結果が変わらないより効率的なSPARQLクエリに書き換える。
- **手法**
SPARQLのグラフパターン内で意味的に冗長な部分をSHACL制約によって排除する。
- **背景**
OPTIONAL演算子はSPARQLにおいて重要だが、複雑になる。
- **結果**
最適なクエリへの書き換えルール提案し、正しいことを証明した。
※この論文で紹介されてるのは一部のみであり、詳細な証明は別の論文。
- **考察**
SHACLによる制約はデータを有用な関係に制限し、より効率的な等価なクエリに書き換えるために使用することができる
- **課題**
これらの書き換えルールをクエリ処理エンジンに統合し、問い合わせ実行を強化することの潜在的な利点を調査する

How is your Knowledge Graph Used: Content-Centric Analysis of SPARQL Query Logs

Luigi Asprino and Miguel Ceriani

- **一言でいうと**
クエリの内容と質的情報に焦点を当ててクエリログを分析するアプローチを提案する
- **動機**
KGがどのように利用されているかを理解し、その長所や短所を評価する
- **手法**
クエリログから推論されるクエリテンプレートを抽出することで、それを要約とする
- **背景**
クエリの構文構造ではなく、内容(特定の語彙など)に焦点をあてた分析は少ない。
- **結果**
抽出されたテンプレートはほとんどのデータセットでログ内の
ユニークなクエリよりも2桁以上少なく、要約の効果があることが
定量的に示された。内容についても定性的に評価した。
- **考察**
個別のクエリ実行だけではなく、短時間に実行された
テンプレートグループを解析することでプロセスの相互作用が
分析できるかも
- **課題**
頻度の高いクエリに焦点を当てたが、頻度が少ないクエリから
得られる洞察について調査

Table 1. Statistics on the LSQ 2.0 dataset before/after summarisation.

Dataset	Execs	Hosts	Queries	H(Q)	Templs.	H(T)	ΔH
Bio2RDF	33 829 184	2 306	1 899 027	15.22	12 296	3.73	11.49
DBpedia	6 999 815	37 056	4 257 903	21.16	17 715	5.58	15.59
DBpedia-2010	518 717	1 649	358 955	17.99	2 223	5.66	12.33
DBpedia-2015/6	6 481 098	35 407	3 903 734	21.01	15 808	5.21	15.80
Wikidata	3 298 254	—	844 260	12.26	167 578	7.47	4.80
LinkedGeoData	501 197	25 431	173 043	14.24	2 748	4.78	9.46
SWDF	1 415 568	921	101 422	14.54	1 826	1.03	13.51

5A: Search, Retrieval and SPARQL II

Industry: Semantic Data and Metadata

東京都市大学 山口研究室

SPARQL edit: Editing RDF Literals in Knowledge Graphs via View-Update Translations

Sascha Meckler and Andreas Harth / Fraunhofer Institute for Integrated Circuits IIS

- **一言でいうと**
非専門家がKGのRDFリテラルを編集できるウェブアプリケーション(SPARQL_edit)の紹介
- **動機**
EKG(Enterprise knowledge graphs)の採用改善のため、非専門家なユーザーでもグラフの値を簡単に維持・更新できるようにする必要性がある
- **リソースの設計方針**
非専門家がクエリ結果を閲覧し、KGの欠いている値を編集・挿入できるWebアプリケーションを作成し、EKGの採用を改善することを目指す
- **新規性**
RDF知識グラフで使用するためのSPARQLベースのビュー更新アプリケーションの実用化
- **リソースの再利用性／活用実績**
RDFフォーマットのビュー構成を簡単に共有できる汎用的なアプローチが可能
- **品質**
変数を2個から30個まで連続的にトリプルパターン数を増やすと、処理時間は直線的に増加
- **リソースの可用性**
オープンソースである(<https://github.com/wintechis/sparqledit>)
- **次にすべきこと**
スケーラビリティとユーザビリティの改善、及び共同作業の可能性の探求に焦点を当てる

Assessing the Generalization Capabilities of Neural Machine Translation Models for SPARQL Query Generation

Samuel Reyd, Amal Zouaq / Polytechnique Montreal

- **一言でいうと**
英語からSPARQLへの翻訳生成モデルの評価
- **動機**
最新のモデルでのシンプルなデータセットのクエリ生成は完ぺきに近くなつたが、著者はこの汎用性と未知の質問クエリ構造に対する能力に疑問を感じている。
- **手法**
英語からSPARQLへ翻訳したペアのデータセットであるLC-QuAD 1.0とLC-QuAd 2.0を用いて精度を評価する。
- **背景**
SPARQLを正式な言語として扱うには少々問題があり、たとえ一つの言葉を似た言葉に置換しても、全く違う結果が出てしまうことがある。
- **結果**
コピーメカニズムを使用すればLC-QuAD 1.0のパフォーマンスは100%近くという結果が出たが、未知の質問が多く含まれているLC-QuAD 2.0では70%程度とまだ課題が残る結果になった。
- **考察**
従来のNMT(Neural Machine Translation)では未知のURIや質問クエリ構造を処理できなかつた。逆に、コピーメカニズムを追加するか、訓練済みモデルを使うことで、未知のURIを含むすべてのテストでパフォーマンスが低下しなかつたので、未知のURIを扱うことができたと言える。
- **課題**
質問の長さ、プレースホルダの数、クエリ内のトリプルの数をより難しいデータに対してどのように一般化するかを示すようなほかの分割基準も検討中。

FedShop: A Benchmark for Testing the Scalability of SPARQL Federation Engines

Minh-Hoang Dang Julien Aimonier-Davat ,Pascal Molli / Nantes Université et al.

- 一言でいうと
SPARQLの連合検索のスケーラビリティをテストするベンチマークとベンチマークを作るツール群を作った
- Resource Type
ベンチマーク
- 動機
既存のベンチマークは、提案されたアプローチやエンジンがフェデレーションメンバーの数が少ない場合、どのように振る舞うかを研究するのに適しているが、数が多くなったときに対応できない。
- リソースの設計方針
 1. テンプレートクエリをプレースホルダーが変数で置き換えられた一般的なクエリに変換する。
 2. これらのクエリを10のベンダーと10のレーティングサイトの構成で実行する。
 3. 実行の結果からプレースホルダーの異なる値の組み合わせをランダムに選択する。
- 新規性
スケーラビリティ実験用の新しいベンチマークであるFedShopを提案
- リソースの再利用性／活用実績
FedShopのデータ生成ツールやベンチマークの設計を用いて、実験の比較ができる
- 品質
BSBMとおなじくe-コマースシナリオに基づいたベンチマークになっている
- リソースの可用性
<https://github.com/GDD-Nantes/FedShop> から利用可能
- 次にすべきこと
ベンチマークにおいて様々なカスタマイズが可能になるようにする。

Facility design metadata as RDF

Dag Hovland, Eirik Nordstrand

- 一言でいうと
RDFを使用して施設設計データのメタデータを効率的に扱うためのデータモデルを提案している
- 動機
スプレッドシートからRDFへの移行に伴うデータの効率的な扱い再利用可能なデータモデルの必要性を感じたから
- 課題
既存の技術では、文書のメタデータを扱うための、効率的で正確かつ再利用可能なデータモデルが提供されていなかった
- 手法
レコードデータモデルとRDFを用いたバージョン管理と出所追跡の最適化
- 評価
データ管理を改善し、設備設計データを表現するためのRDFの導入を促進できる
- 考察
データ管理の改善とその手法の適応性を通じて、設備設計データの表現における革新的な解決策を提供した
- 知見(今後の課題)
提案されたアプローチの採用と可用性への抵抗を克服し、さらに多様な産業データの表現にRDFを適用すること

Linked data supporting the legislative decision process

Johan Delaure / redpencil.io and Alvin Demeyer /Flemish government

- **一言でいうと**
法律にリンクデータの概念と構造を適用し、そのデジタル変革を促進することが成功の鍵である。
- **動機**
法律はリンクデータ形式を活用することで、法的文書の意味豊かな説明を実現し、法令の前段階の意味的サポートを目指している。これはフレミッシュ政府の semantic.works プラットフォームとビジネスで承認されたデータモデルによって実現されている。
- **課題**
意思決定のプロセスをデジタル変革する際に、アジャイルな開発が必要であり、新たな要件や変更に柔軟に対応する必要がある。これは伝統的なデータベースに基づくアプローチでは難しく、ビジネス承認データモデルが一貫性を確保する鍵となっている。
- **手法**
semantic.works プラットフォームを使用し、ビジネスで承認されたデータモデルを基に、法令制定のプロセスをサポート。セマンティックデータモデルを活用し、データのリンク付けと共有が、政治的な意思決定の複雑なプロセスをサポートする鍵となっている。
- **評価**
プラットフォームは4年間で数百のアジェンダを処理し、文書の公開などの複雑なタスクを成功裏に実行しており、法令の増加に追いつく助けとなっている。ユーザーの数や公開された法的文書のページ数などが指標となっている。
- **考察**
デジタル変革において、アジャイル開発とビジネス承認データモデルが成功の鍵であり、政治的な意思決定プロセスは継続的な進化と新たな洞察への対応が求められる。リンクデータ構造はドメインのモデリングだけでなく、プロセスの維持においても主要な手段となっている。
- **知見**
異なる行政機関がデジタルプロセスに参加するためには、まだ多くの努力が必要であり、これは伝統的なプロセスからクラウド共有情報ウェブへの転換を意味している。これには法令の増加や複雑さ、文書の大きさ、処理速度の要求が促進要因となっている。

Semantic Cloud System for Scaling Data Science Solutions for Welding at Bosch

Zhuoxun Zheng/Bosch Center for AI, Baifan Zhou/University of Oslo, Zhipeng Tan/RWTH Aachen University, et al.

- 一言でいう
クラウドの専門家ではないユーザー向けのクラウドシステム
- 動機
Society4.0に伴い、自動化のためIoT技術に依存するスマート工場と呼ばれる工場では、大量のデータを扱うためクラウド技術の需要が高まっているが、このユーザーはクラウドの専門家ではないことが多い。
- 課題
クラウド上にソリューションを開発したりする場合はクラウドの専門家による支援が必要であり、コストとメリットのバランスを考慮した計画が必要となる。
- 手法
ETL(Extract, Transform, Load)を取得、スライス、準備、保存という4つのステップからなるパイプラインに分解することでETLの最適化されたクラウドを自動的に実現するSemCloudを提案している。
- 評価
SemCloudを使用すると、BoschのセマンティックETLは少なくとも2倍速くなり、クラウド設定の最適化時間は1.12秒に短縮された。
- 考察
SemCloudは、より多くのユーザーがクラウドシステムを使用できるようにし、人材トレーニングとデータ処理にかかる時間とコストを大幅に削減し、Boschのデータサイエンスソリューションに利益をもたらした。
- 知見
記載なし。

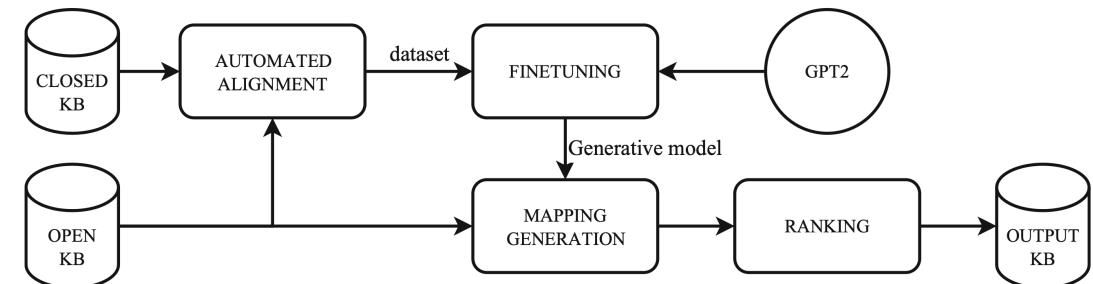
2B: Knowledge Extraction

Wataru Hirota (Stockmark)

Mapping and Cleaning Open Commonsense Knowledge Bases with Generative Translation

Julien Romero/IPParis, Simon Razniewski/Bosch Center for AI

- 一言でいうと
OpenIE の結果を既存の KG にアラインする方法を提案
- 動機
OpenIE で得られた知識はエンティティ・関係が正規化されておらず使いにくい
- 手法
OpenIE の結果を事前に定義されたスキーマにあうように出力する
翻訳モデルを学習し、使用する
- 背景
生成言語モデルの高い翻訳精度を活かすことができる
- 結果
ground-truth の KG と比較した recall / precision がルールマイニング手法よりも高かった
- 考察
ルールベース手法と異なり、生成モデルの手法は関係に応じて柔軟にエンティティを変えられたのが精度向上に繋がった
- 課題
今回使用したのが GPT-2 で、まだ最近の LLMs に対して検証ができていない



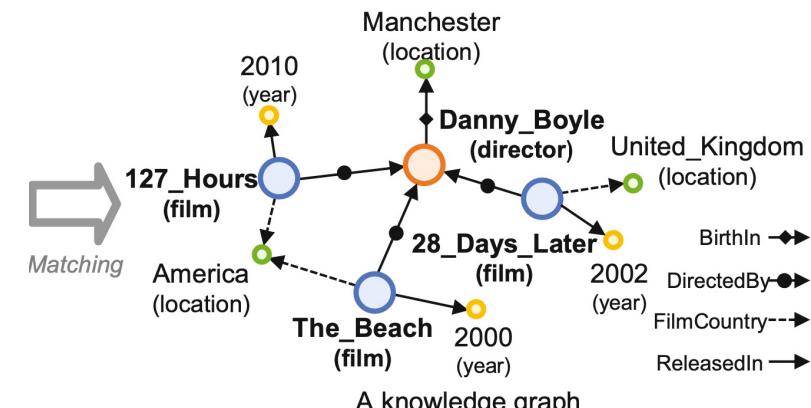
Dependency-Aware Core Column Discovery for Table Understanding

Jingyi Qiu, Aibo Song, Jiahui Jin, Xiaolin Fang, Jingyi Ding, Tianbo Zhang/Southeast University,
Jianguo Qian/State Grid Zhejiang Electric Power Company

- 一言でいうと
リレーションナルテーブルから (1) entity のコア概念を表す "core columns" と (2) 列間の依存関係を自動的に特定する方法を提案
- 動機
テーブルデータから KG を作るためには core columns や依存関係の発見が重要
- 手法
「列 y の値が列 x に依存している」度合いを表す $\text{dep}(x, y)$ を定義し, iterative に計算
その結果できた列の依存関係ツリーの一番根元の依存元の集合を core columns とする
- 背景
今まで core columns の発見に簡単なヒューリスティックしか用いられず、
"id" など無意味な列を core columns とみなしがちだった
- 結果
人手でアノテートした core columns との一致率が既存のヒューリスティックより高かった
- 考察
提案手法は iterative な方法を採用しており、その停止条件の適切な設定により
core columns の発見精度が向上することがわかった
- 課題
特に言及無し

Core Columns					
id	filmcountry	releasetime	name	director	location
1	america	2010	127 hours	danny boyle	Manchester
2	united kingdom	2002	28 days later	danny boyle	manchester
3	united states	2000	the beach	danny boyle	Manchester

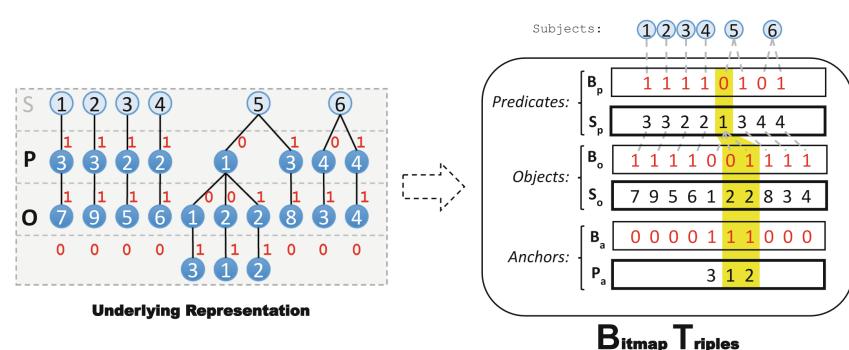
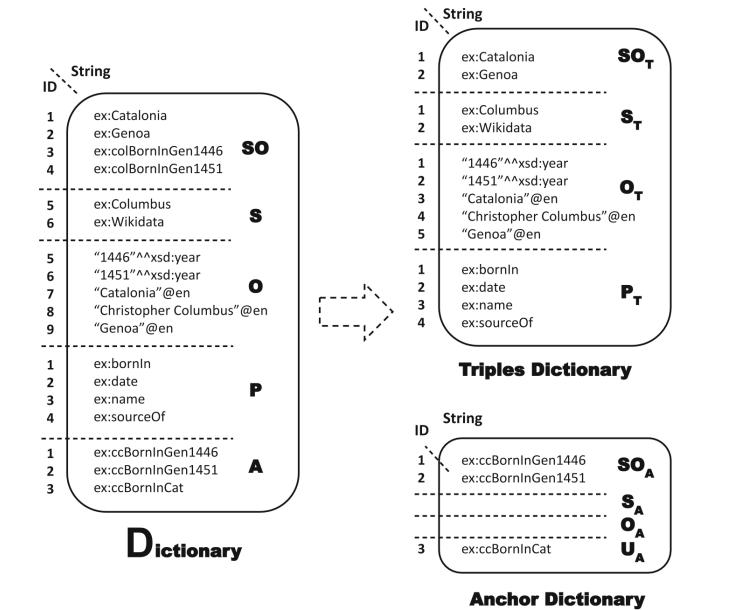
A table about movies and directors



Compact Encoding of Reified Triples using HDTr

Jose M. Gimenez-Garcia/Univ. of Valladolid, Thomas Gautrais/Univ. de Lyon, Javier D. Fernández/Data Science Acceleration, Miguel A. Martínez-Prieto/Univ. of Valladolid

- 一言でいうと
reification を Header-Dictionary-Triples (HDT) 互換の形式で表現する HDTr を提案
- 動機**
KG の triplet そのものの情報 (context information, 情報ソースなど) を付与する方法の
1つに reification がある。これの省データスペースなデータ構造を実現したい
- 手法**
reification の anchor (各 statement に付与される仮想ノード) を
HDT dictionary に導入する
- 背景**
reification は anchor ノードが増えるため、データ量が大きくなりがちだった
- 結果**
RDF reification や NdFluents、素の HDT に比べデータの圧縮率と検索速度が向上
- 考察**
context information が少ない方がより HDTr の圧縮率が多手法に比べ高くなった
- 課題**
より efficient な reification の表現方法を追求したい



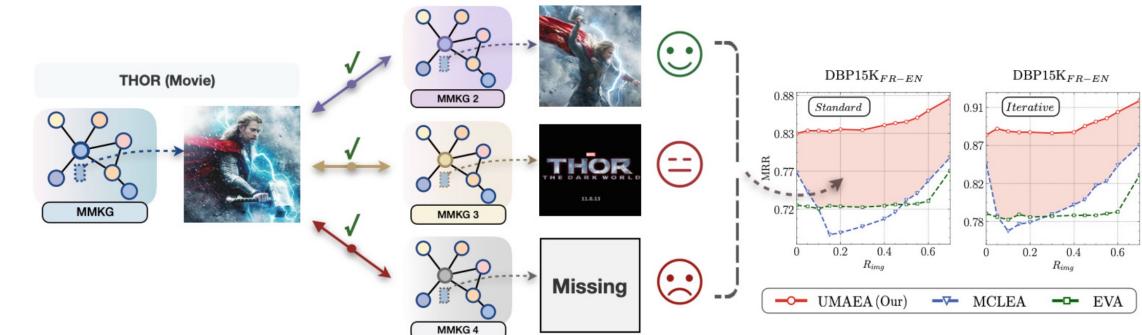
9B: Entity Alignment

Wataru Hirota (Stockmark)

Rethinking Uncertain Missing and Ambiguous Visual Modality in Multi-Modal Entity Alignment

Zhuo Chen/Zhejiang Univ., Lingbing Guo, Yin Fang, Yichi Zhang, Jiaoyan Chen, Wen Zhang, Jeff Z. Pan, Yangning Li and Huajun Chen

- 一言でいうと
Multi-Modal Entity Alignmentにおいて、画像が曖昧な場合・不十分な場合でも安定した精度が出る手法を提案
- 動機
実正解の Multi-Modal Entity Alignment はしばしば画像に欠損があつたり不十分
- 手法
画像の incompleteness, noise の影響を緩和するような
ニューラルネットワークの損失関数を導入する
- 背景
画像の incompleteness, noise を緩和する機構がない場合、
画像が EA の精度にネガティブな影響を及ぼす
- 結果
DBpedia のいくつかの言語対 (ZE-EN, JA-EN, FR-EN) で実験し、いずれのデータに対しても既存手法より高い H@1, MRR を記録
- 考察
entity の画像の含有率が低い場合、特に提案手法の優位性が目立った
- 課題
画像以外の modality (例: attribute) の影響はまだ調べられていない



Iterative Geographic Entity Alignment with Cross-Attention

Alishiba Dsouza/Univ. of Bonn (1), Ran Yu/(1), Lamarr Institute for Machine Learning and Artificial Intelligence (2),
Moritz Windoffer (1) and Elena Demidova (1,2)

- 一言でいうと
Geographic entity (例: ベルリン) に対する entity alignment 手法を提案

- 動機**

Wikidataなどの汎用的な知識グラフには有名な geographic entity しか含まれていないため、
OpenStreetMapのような community ベースで更新されている geographic entity を取り込みたい

- 手法**

EA を class alignment (属性間のアライメント) と entity alignment の2ステップに分け iterative に解く

- 背景**

geographic entity は属性が sparse でヘテロ (スキーマがばらばら) であるので、事前に与えられた教師データ以外も利用できる iterative な手法が良いのではと考えた

- 結果**

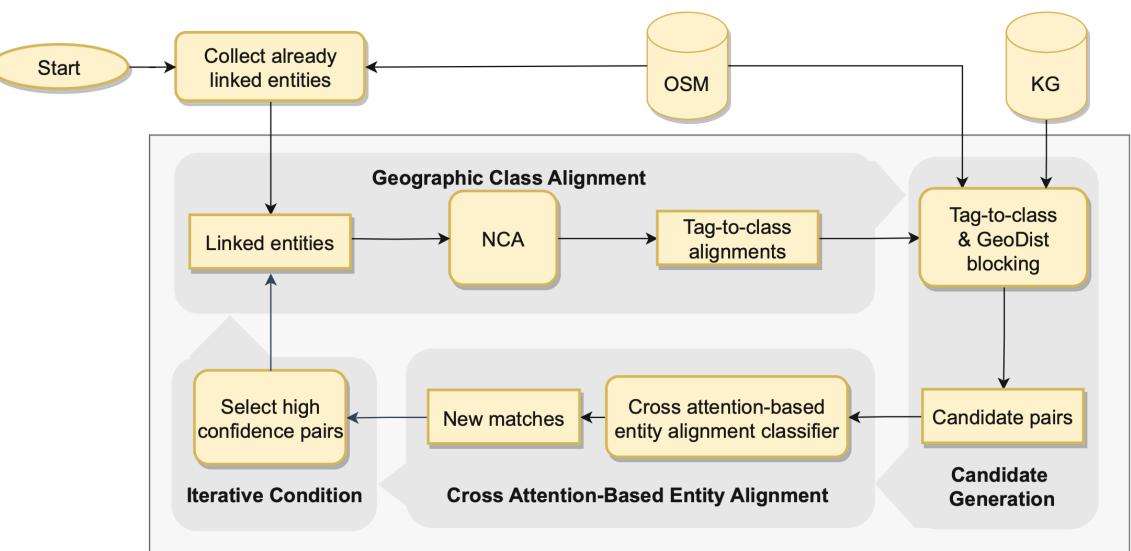
OSM と Wikidata のアライメントで、既存手法を上回る F1 値を記録

- 考察**

アメリカなどデータの量が多い国は既存手法の精度も高かったが、
提案手法はフランスやドイツなどデータが少ない国でも精度が高くなった

- 課題**

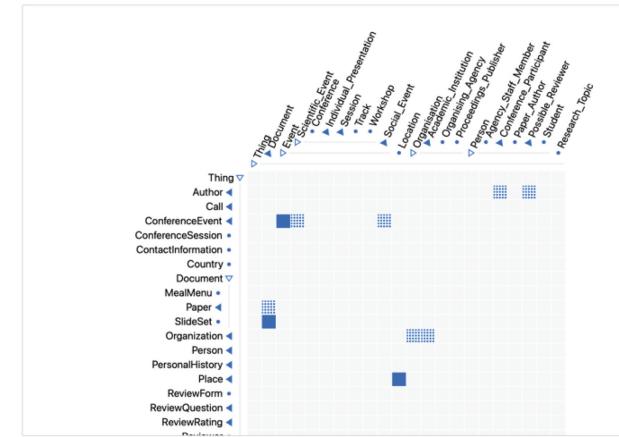
とくに言及無し



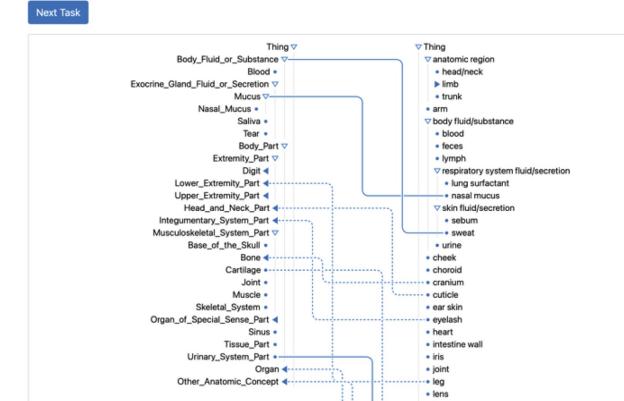
Visualizing Mappings Between Pairwise Ontologies - An Empirical Study of Matrix and Linked Indented List in Their User Support During Class Mapping Evaluation

Bo Fu, Allison Austin and Max Garcia/California State University Long Beach

- 一言でいうと
人手によるオントロジーのマッピング作業・可視化についてのユーザースタディー
- 動機
オントロジーの可視化についての研究は多くあるが、オントロジーのペアの関係をどう可視化するかについての研究は少ない
- 手法
被験者にオントロジーのペアに関する質問（例：source の “Author” に対応する target のクラスの数は？）を各ツールを使って解いてもらい、その正答率と所要時間を比較する
- 背景
オントロジーマッチングで特に重要な操作について比較したい
- 結果
LIL（図下側）を使った被験者は matrix（図上側）の被験者に比べ質問の正答率が高く、所要時間が短い傾向にあった。
また LIL は目の物理的な移動距離が小さかった。
- 考察
“学術会議”など被験者にとってあまりなじみのないオントロジーの場合、両可視化による差はあまり見られなかった
- 課題
クラス同士だけではなく、instance や property の比較におけるインタラクションも研究したい



(a) Mapping Visualized in A Matrix in the Conference Domain



(b) Mapping Visualized in A Linked Indented List in the Anatomy Domain

2A: Knowledge Engineering with Large Language Models

大阪電気通信大学 古崎研究室 (脇所, 鈴木, 福田)

[著者名] Yiming Tan, Dehai Min, Yu Li, Wenbo Li, Nan Hu, Yongrui Chen, and Guilin Qi
 / Southeast University, Anhui University, Southeast University, Ministry of Education

表1 全体的な評価結果

- 一言でいうと ChatGPTと従来のKBQA モデルの比較とフレームワークを紹介。

• 動機

不明。

• 手法

ChatGPTと従来のKBQAモデルを使用するデータセットを変えつつ比較している。

• 背景

ChatGPT が従来のものに取って代わることができるか調べることの関心が高まっていること。

• 結果 右に示す。

• 考察

現在の SOTA の従来のモデルは 8つのテストセットのうち4つで、その後はリリースされた GPT-4 は4つのテストセットで最も優れていた。GPT-4 モデルと SOTA モデルを比較すると、以下のことを発見した。

- GPT Familyによって、そのゼロショット能力は近づいており、従来の深層学習および知識表現モデルを超えていている。
- GPT Familyのモデルを比較すると、全てのデータセットにおいて、新しいモデルの方がパフォーマンスが優れている。

• 課題

オープンドメイン KBQA から得られた結論がどのようなものであるかを検証するための普遍的ドメインと固有のドメインに関して調べること。さまざまなタイプのテストの追加。

Datasets	KQAPro	LC-quad2	WQSP	CWQ	GrailQA	GraphQ	QALD-9	MKQA
	Acc	F1	Acc	Acc	Acc	F1	F1	Acc
SOTA(FT)	93.85 [29]	33.10 [31]	73.10 [15]	72.20 [15]	76.31 ‡	31.8 [13]	67.82 [32]	46.00 [22]
SOTA(ZS)	94.20 [25]	-	62.98 [50]	-	-	-	-	-
FLAN-T5	37.27	30.14	59.87	46.69	29.02	32.27	30.17	20.17
GPT-3	38.28	33.04	67.68	51.77	27.58	38.32	38.54	26.97
GPT-3.5v2	38.01	33.77	72.34	53.96	30.50	40.85	44.96	30.14
GPT-3.5v3	40.35	39.04	79.60	57.54	35.43	47.95	46.19	39.05
ChatGPT	47.93	42.76	83.70	64.02	46.77	53.10	45.71	44.30
GPT-4	57.20	54.95	90.45	71.00	51.40	63.20	57.20	59.20

縦の比較は、同一データセットにおけるモデルの比較。現在の SOTA の伝統的な KBQA モデル (ファインチューニング (FT) およびゼロショット (ZS))、GPT LLM Family、および非GPT LLMが比較対象。
 データセットのうちGraphQ、QALD-9、LC-quad2では、使用される評価指標は F1。他のデータセットは精度 (完全一致) を使用。

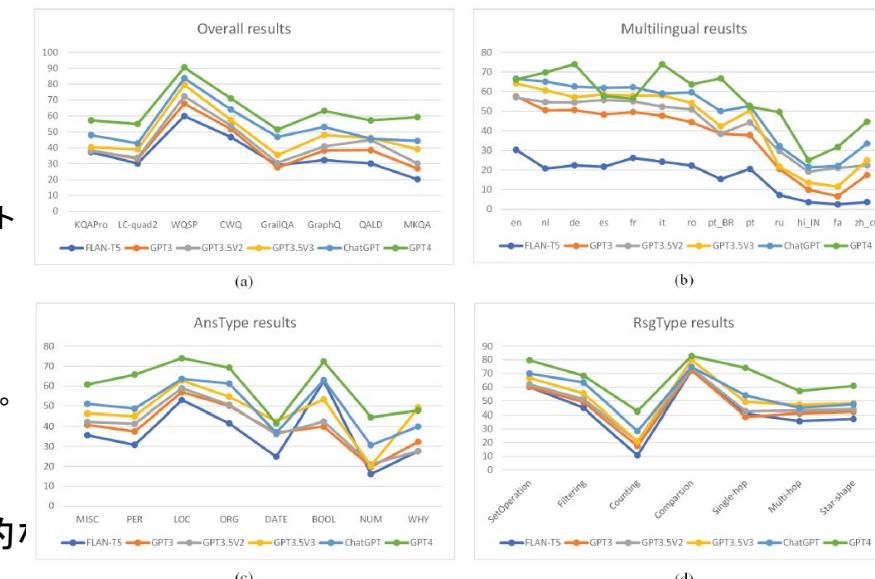


図1 (a)全体的な結果 (b)言語 (c)Ansタイプ (d)Rsgタイプによる各データモデルセットのパフォーマンスの折れ線グラフ

LLMs4OL: Large Language Models for Ontology Learning

Hamed Babaei Giglou(B) , Jennifer D'Souza , and Sren Aöuer/TIB Leibniz Information Centre for Science and Technology, Hannover, Germany

- **一言でいうと**
オントロジー学習(OL)のための大規模言語モデル(LLM)を利用するLLMs4OL手法を提案している。
- **動機**
LLMは、その言語パターン捕捉能力を用いて、自然言語テキストから知識を自動的に抽出し、構造化するOLに効果的に適応できるのか？という仮説の検証をするため。
- **手法**
9つのLLMに対して、Zero-Shot プロンプティングを行い、「用語の分類・分類の発見・非分類関係の抽出」という主要なOLタスクに関して評価する。
- **背景**
LLMは自然言語処理において大きな進歩を遂げており、様々な知識領域における複雑な言語パターンを捉える能力が示されている。
- **結果**
基礎的なLLMは、高度な推論スキルやドメイン知識を必要とするオントロジーの構築には十分に適していないことが分かった。
- **考察**
LLMを微調整することで知識獲得のボトルネックを軽減し、オントロジー構築をアシストできる可能性がある。
- **課題**
OLタスクに特化してLLMを強化すること・評価を多様な知識ドメインに拡大することなどが挙げられている

Text2KBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text

Nandana Mihindukulasooriya 1, Sanju Tiwari 2, Carlos F. Enguix 2, and Kusum Lata 3 / 1 IBM Research Europe, Dublin, Ireland, 2 Universidad Autonoma de Tamaulipas, Victoria, Mexico, 3 Sharda University, Greater Noida, India

- **一言でいうと**
任意のオントロジーに基づいたテキストからKGを生成するLLMの能力を測定するベンチマーク
- **Resource Type**
Evaluation Benchmark
- **動機**
LLMとKGの併用に関する研究のための、ベンチマークの確立
- **リソースの設計方針**
正しい関係を用いて事実を抽出する能力を評価
RDF/OWL表現の処理や生成の能力を評価は対象外
- **新規性**
テキストからのKG生成タスクにおける初のベンチマーク
- **リソースの再利用性／活用実績**
著者らが主催するテキストからのKG生成のワークショップで提供予定
- **品質**
不明
- **リソースの可用性**
データセットはzenodo, コードはGithubで入手可能
- **次にすべきこと**
性別や人種などのバイアスを考慮しベンチマークを拡張

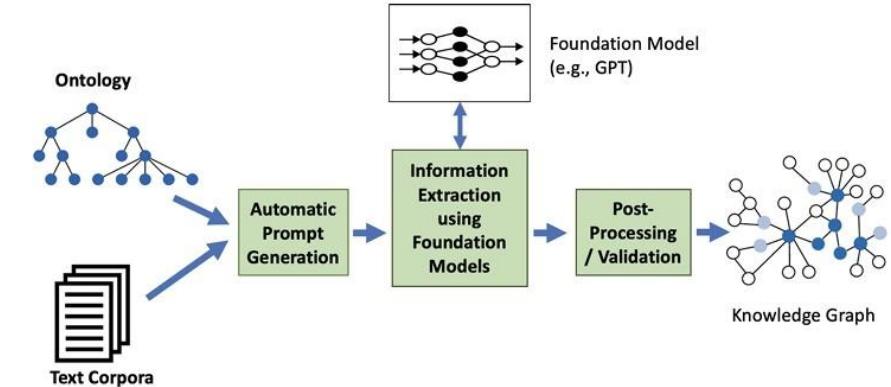


Fig. 1. Generating knowledge graphs from text guided by ontologies

Table 2. This table summarizes average evaluation metrics for all ontologies in Wikidata-TekGen and the DBpedia-WebNLG datasets.

Dataset	Model	Variant ^a	Fact Extraction			OC	Hallucinations		
			P	R	F1		SH	RH	OH
Wikidata-TekGen	Vicuna	All	0.38	0.34	0.35	0.83	0.17	0.17	0.17
		Selected	0.42	0.39	0.38	0.84	0.11	0.16	0.14
		Unseen	0.32	0.32	0.32	0.86	0.07	0.14	0.14
	Alapaca LoRA	All	0.32	0.26	0.27	0.87	0.18	0.13	0.17
		Selected	0.33	0.27	0.28	0.87	0.12	0.13	0.17
		Unseen	0.22	0.22	0.22	0.86	0.09	0.14	0.26
DBpedia-WebNLG	Vicuna		0.34	0.27	0.30	0.93	0.12	0.07	0.28
	Alpaca-LoRA		0.32	0.23	0.25	0.91	0.16	0.09	0.38

^aRefer to Sect. 3 for details.

10A: Ontology engineering and ontology patterns

大阪電気通信大学 古崎研究室（浦中，古崎，松村）

The Wikibase Approach to the Enslaved.Org Hub Knowledge Graph

Cogan Shimizu, Pascal Hitzler, Selia Gonzalez-Estrecha, Jeff Goeke-Smith, Dean Rehberger, Catherine Foley, and Alicia Sheill/Wright State University , Dayton , Ohio , USA, Kansas State University , Manhattan , USA, Michigan State university , East Lansing , USA

- 一言でいうと

歴史的な奴隸貿易を記録するための知識グラフ(KG)である
Enslaved.org HUBの設計、実装、展開に使用された方法論について。

- 動機

KGを作成、展開、定義するためのプラットフォームは数多く存在し、
Wikibase・プラットフォームの使用は利点が多くある。

- 手法

スキーマの開発と転送、KGの具体化、検証、相互参照の解決、
Wikibase・プラットフォームへのKGの展開。(Fig.2)

- 背景

KGがどのように利用されるかは、利用者の興味、知識、経験に関わる。

- 結果

KGのスキーマの開発、スキーマの調整、データの重複排除、KGの展開。

- 利用実績

過去 1 年間で月間350万人のユニークユーザー。

- 学んだ教訓やベストプラクティス

Wikibase・プラットフォームの導入は、技術的習熟度が低いコミュニティにとって有益である。

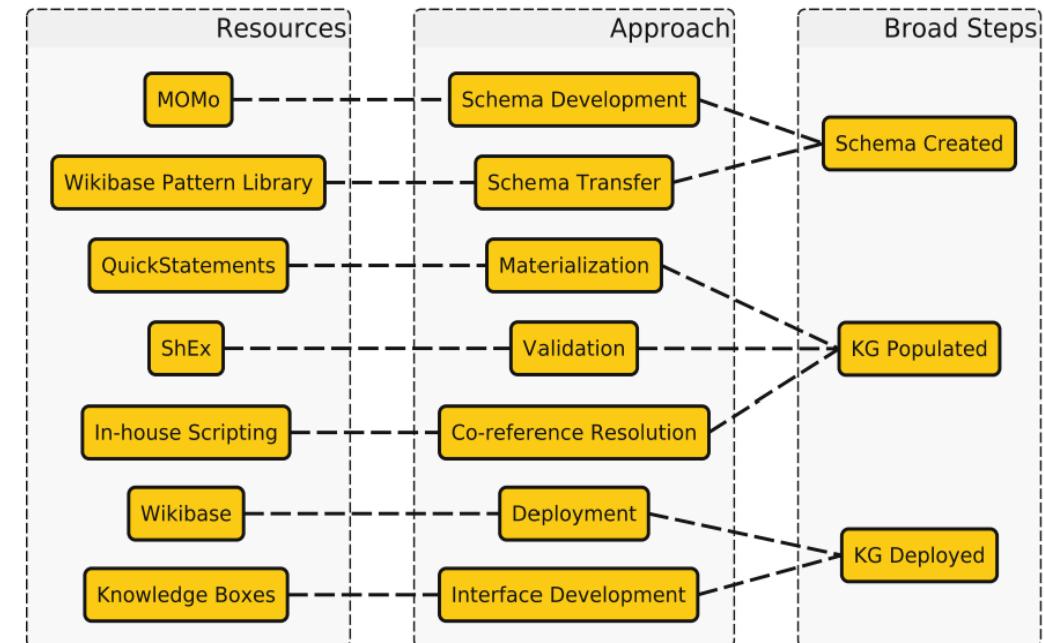


Fig. 2. This diagram displays the resources used (left) to implement our approach (middle) for creating and deploying the Enslaved.Org Hub knowledge graph. These steps are broadly categorized (right) for simplicity.

The RML Ontology: A Community-Driven Modular Redesign After a Decade Resource Track

of Experience in Mapping Heterogeneous Data to RDF

Ana Iglesias-Molina¹, Dylan Van Assche², Julian Arenas-Guerrero¹, Ben De Meester², Christophe Debruyne³, Samaneh Jozashoori^{4,5}, Pano Maria⁶, Franck Michel⁷, David Chaves-Fraga^{1,8,9}, and Anastasia Dimou⁹/ ¹Universidad Politecnica de Madrid, ²Ghent University-imec, ³University of Liege
⁴metaphacts GmbH, ⁵Leibniz Information Center for Science and Technology, ⁶Skemu, ⁷University CotedAzur, ⁸Universidade de Santiago de Compostela, ⁹KU Leuven

- 一言でいうと

RDF Mapping Language (RML)オントロジーを構築し、関係データベースからRDFグラフを構築するためのマッピングルールの定義、およびRMLに準拠したシステム開発を支援する。5つのモジュールから構成されるオントロジーとなっている。記述のチェックにはSHACLが利用される。

- Resource Type

Ontology

- 動機

既存のR2RML(Relational to RDF Mapping Language)では、十分に対応できていない事項が多くあり、それらをカバーしたRMLを構築する必要がある。

- リソースの設計方針

Linked Open Terms (LOT) methodologyによる4つのステージ（要求仕様の策定、実装、公開、メンテナンス）に沿って構築。

- 新規性

スキーマ・データの変換、RDFのCollections and Containers、RDF-starなど、これまでのR2RMLで扱えなかった点に対応。

- リソースの再利用性／活用実績

既にいくつか実プロジェクトにおいてRMLオントロジーがサポートされている。

- 品質

OOPS! によるオントロジーのエラーチェック、および HermiT reasonerを用いた整合性検証を実施。

- リソースの可用性

GitHubレポジトリでCC-BY4.0にて公開するとともに、issueを管理。W3C KG Construction Community Groupにて議論。

- 次にすべきこと

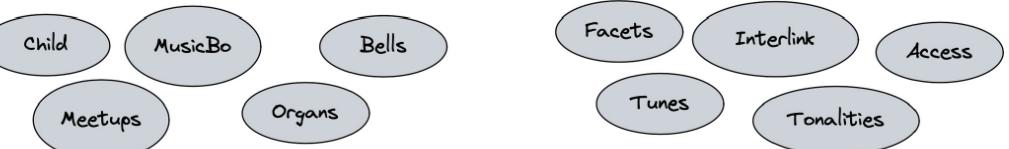
W3C Recommendationに向けた議論の継続。SHACL shapesによる検証（validation）についても改良を予定。

The Polifonia Ontology Network: Building a Semantic Backbone for Musical Heritage

Jacopo de Berardinis, Valentina Anita Carriero, Nitisha Jain, Nicolas Lazzari, Albert Meroño-Peñuela, Andrea Poltronieri and Valentina Presutti
/ King's College London, University of Bologna

- 一言でいうと

音楽データ、表現、歴史的情報源、楽器の4つのモジュールを中心として
作成する音楽文化遺産のためにThe Polifonia Ontology Networkの提案



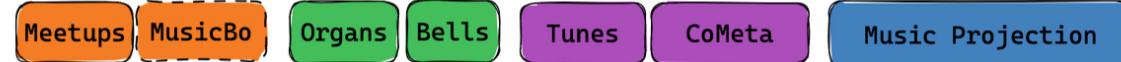
- 動機

様々な言語で作成された音楽、具体的な場所と結びついた音楽コンテ
ンツ等の多様性あふれる音楽遺産の研究、保存



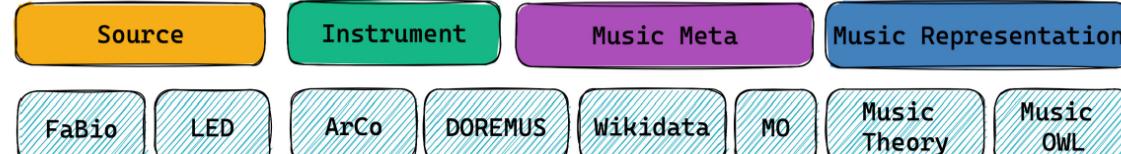
- 手法

オントロジーエンジニアリング



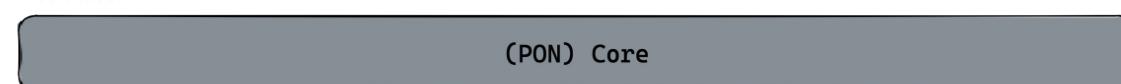
- 背景

膨大にある音楽遺産に関する資料の保存



- 結果

15個のオントロジーと361個のコンピテンシー質問からなるデータの公開



- 利用実績

61人へのオンライン調査による音楽オントロジーの背景、関連性、使用に関する質問

- 課題

既存の音楽オントロジーの多くが独自に作成されたものであり再利用や拡張が困難

1A: Ontologies and Knowledge Graphs I

7A: Ontologies and Knowledge Graphs III

青山学院大学 森田研究室

Disentangled Contrastive Learning for Knowledge-aware Recommender System

Shuhua Huang¹, Chenhao Hu¹, Weiyang Kong¹, and Yubao Liu^{1,2}

¹Sun Yat-Sen University, Guangzhou, China, ²Guangdong Key Laboratory of Big Data Analysis and Processing, Guangzhou, China

- 一言でいうと

知識を意識したレコメンデーションのための、解きほぐされた対比学習フレームワークを構築する新しいモデル DCLKR の提案

- 動機

アイテム属性のさまざまな側面に基づいて多面的なユーザーの好みを探ることの重要性を強調し、解きほぐされた表現学習によって多面的なユーザーの好みをモデル化するというアイデア

- 手法

アイテムナレッジグラフを複数の側面に分解し、インタラクショングラフをエンコードし、2つのビュー間で整列された項目表現に対してビュー間対比学習を行う。

- 背景

各ユーザーと各アイテムの単一の表現しか学習しないため、アイテム属性のさまざまな側面に基づいて多面的なユーザーの好みを発見するには不十分

- 結果

3つのベンチマークデータセットで広範な実験を実施し、アブレーション研究で各コンポーネントの有効性を示した

- 考察

アイテム属性のさまざまな側面に基づいてユーザーの多面的な好みを捕捉することができ、知識ビューと協調ビューの両方からの特徴が保存されるため、表現が DCLKR にとってより有益なものになる

- 課題

- 記載なし

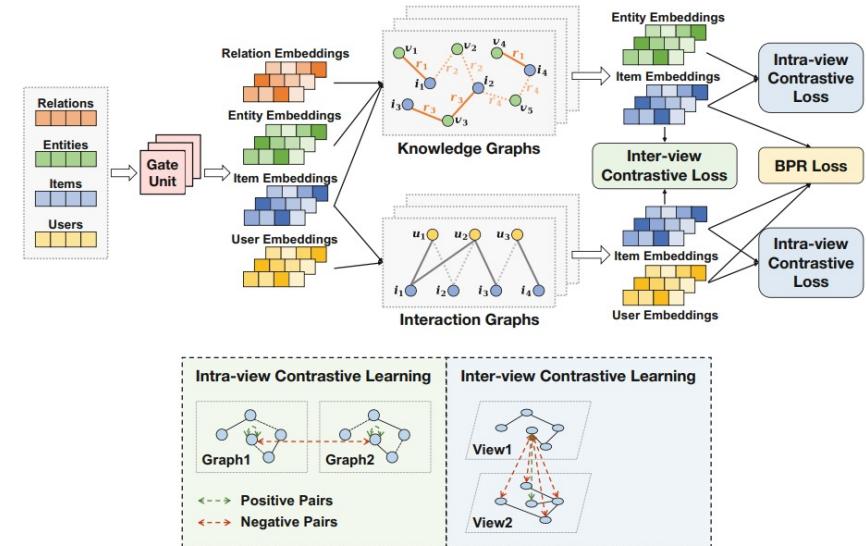


Fig. 2. Illustration of the proposed DCLKR model. The upper part is the model framework of DCLKR, and the lower part is the details of intra-view and inter-view contrastive learning mechanism. Best viewed in color.

SemOpenAlex: The Scientific Landscape in 26 Billion RDF Triples

Michael Färber¹, David Lamprecht¹, Johan Krause¹, Linn Aung², Peter Haase² /

¹ Institute AIFB, Karlsruhe Institute of Technology (KIT), ² metaphacts GmbH, Walldorf

- 一言でいうと
あらゆる学術分野にわたる広範な学術データを網羅した、約260億のRDFトリプルを持つ大規模なデータセット（知識グラフ）

• Resource Type

Dataset

• 動機

研究者の広範かつオープンなデータにアクセスする必要性の高まり

• リソースの設計方針

AWS S3 から OpenAlex のスナップショット (JSON) をRDF形式に変換

• 新規性

広範なメタデータ / SPARQL対応 / 埋め込みの提供 / 半自動の月次更新

• リソースの再利用性／活用実績

研究者や機関をランキングし、研究の動向を分析可能

• 品質

13のエンティティタイプと87の関係タイプを含む

• リソースの可用性

CC0ライセンスで提供され、利用者が自由にビルドアップして拡張・再利用可能

• 次にすべきこと

資金プログラムに関するメタデータを組み込むことで、政府や機関の資金配分に関する詳細で包括的な評価を可能にする

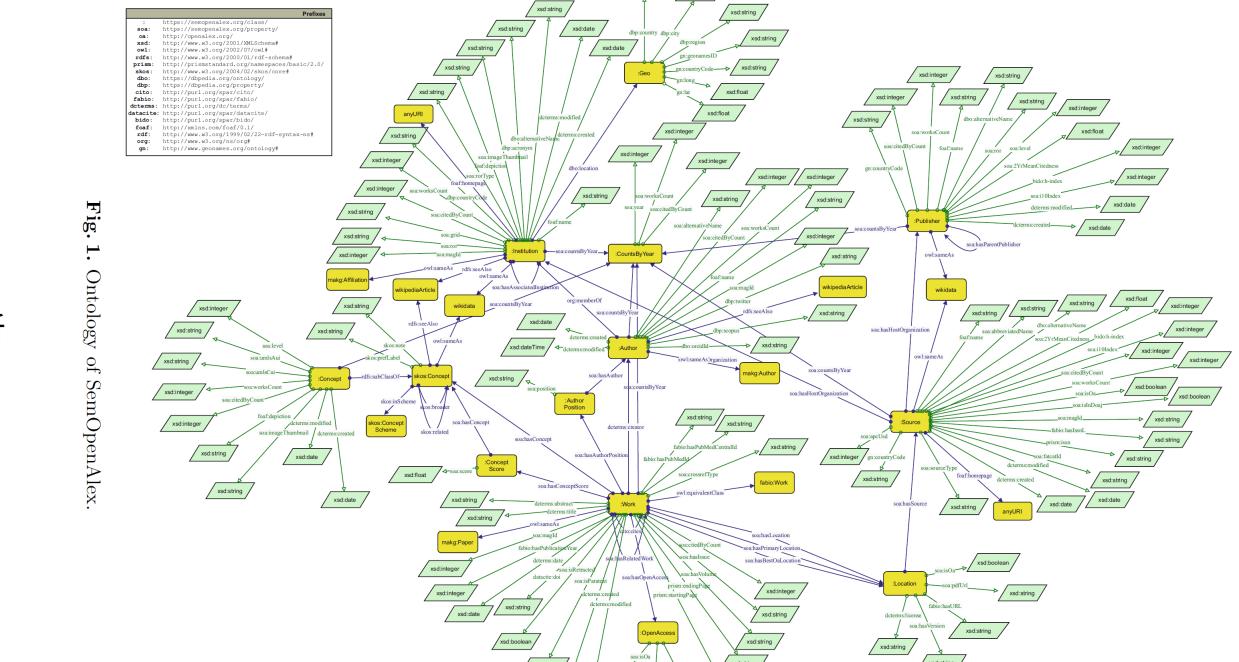


Fig. 1. Ontology of SemOpenAlex.

The World Literature Knowledge Graph

Marco Antonio Stranisci^{1(B)}, Eleonora Bernasconi², Viviana Patti¹, Stefano Ferilli², Miguel Ceriani^{2,3}, and Rossana Damiano¹

- **一言でいうと**

国籍、人種、性別に関する偏見に対処するために開発された文学知識ベースの提案。

- **動機**

文学作品を通じて世界の多様な文化や社会を理解するための新しい視点を提供。

- **手法**

- 作家の民族的な表現の不足をモデル化し、異なるオントロジーを統一されたデータモデルにマッピング。

- SKATEBOARDに統合され、つながりを視覚的に表現。

- **背景**

近年、文学作品を通じて異なる文化や時代の理解を深めるためのセマンティックリソースの開発や、リンクトデータの視覚化プラットフォームの設計への注目。

- **結果**

トピック、著者、作品間の関連性を探索し、視覚化するための更新されたカスタマイズ可能な知識グラフを構築し、多様な研究分野での応用が期待される。

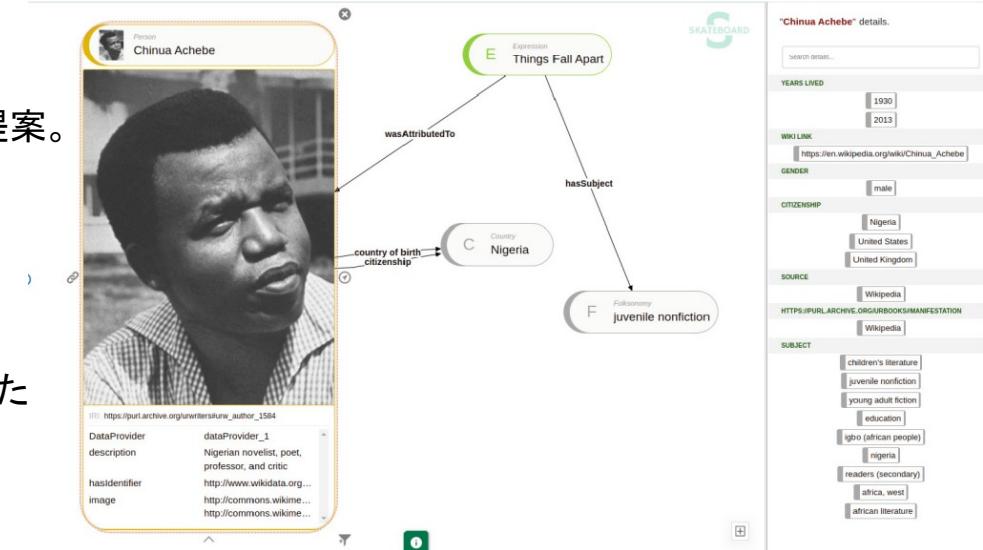
- **考察**

新しい作家を発見するための従来の文学検索ツールに対する有効な代替手段となり得る。

- **課題**

非専門家ユーザーにとってのユーザーエクスペリエンスの改善が課題。

公平な推薦を提供するための影響を評価する目的で、知識グラフに基づく推薦システムのテストを行う予定。



The Holocaust Archival Material Knowledge Graph

Herminio García-González¹ and Mike Bryant^{2,3} / 1 Kazerne Dossin, Goswin de Stassartstraat 153, 2800 Mechelen, Belgium

2 King's College London, Strand, London WC2R 2LS, UK

3 NIOD Institute for War, Holocaust and Genocide Studies, Herengracht 380, 1016 CJ Amsterdam, Netherlands

- **一言でいうと**

Holocaust研究のためのアーカイブメタデータをLODとして提供する取り組み.

※ Holocaustとは、第二次世界大戦中にナチ党支配下のドイツ国やその占領地においてユダヤ人などに対して行った大量虐殺を指す.

- **動機**

Holocaust研究のための資料をよりアクセスしやすく統合された形で提供したい.

- **手法**

1. EHRIポータルのデータは、国、アーカイブ機関、アーカイブ記述の3つのエンティティを基盤としており、これらの関連性をモデル化する。
2. 国際アーカイブ評議会(ICA)の概念に沿ってEHRIのデータを整合し、RDF形式でモデリングされる。
3. 収集されたデータは、ShExML言語を使用してマッピングされ、Turtleファイルに変換される。最終的にナレッジグラフを形成する。

- **背景**

Holocaustに関する資料は、第二次世界大戦による人々や行政機関の移動、証拠の意図的な破壊により、情報源が断片化され分散しているため、情報を収集する際に課題が生じる。

- **結果・利用実績**

- Holocaust関連資料の大規模なナレッジグラフが構築された。これには6,571,095のトリプルが含まれる。
- データセットは DBpediaやCDECの人物データベースなど他の知識ベースとのリンクによって、ユーザの探索が容易になった。

- **学んだ教訓やベストプラクティス**

- セマンティックウェブ技術を活用することで、データのアクセス性を向上させた。
- データを他の知識ベースとリンクすることで、データの価値を高め、相互運用性を向上させた。

MMpedia: A Large-scale Multi-modal Knowledge Graph

Yinan Wu, Xiaowei Wu, Junwen L and Yue Zhang / School of Information Science and Engineering

Haofen Wang / College of Design and Innovation,

WenDu and Zhidong He / DS Information Technology

Jingping Liu and Tong Ruan / School of Information Science and Engineering

- 一言でいうと
大規模なマルチモーダルKGについて
- Resource Type
Knowledge Graph
- 動機
既存のマルチモーダルKGはエンティティ数が少なく拡張が困難
- リソースの設計方針
ウェブ検索画像に対して複数ステップのフィルタリングを行う
- 新規性
既存のマルチモーダルKGより大規模
- リソースの再現性／活用実績
Wikidataのタイプ情報を活用し画像フィルタリング
- 品質
3人の被験者による正誤判定の結果、正答率は81.14%
- リソースの可用性
Webページ (<https://github.com/Delicate2000/MMpedia>) で公開
- 次にすべきこと
様々な下流タスクへの応用と評価

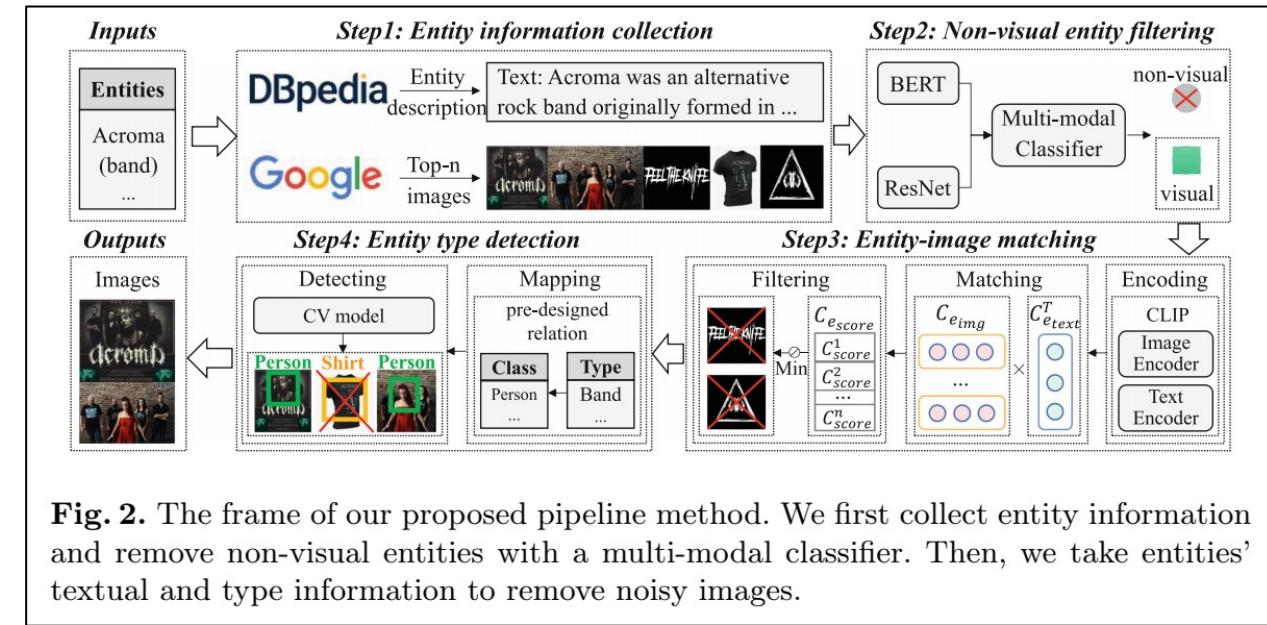


Fig. 2. The frame of our proposed pipeline method. We first collect entity information and remove non-visual entities with a multi-modal classifier. Then, we take entities' textual and type information to remove noisy images.

6B: Ontologies and Knowledge Graphs II

野本昌子（理化学研究所）

HOLY: An Ontology covering the Hydrogen Market

Kiara M. Ascencion Arevalo, Christoph Neunsinger, Roland Zimmermann, Ralph Blum and Kendra Weakly/Technische Hochschule Georg Simon Ohm

- 一言でいうと
水素エネルギー市場をモデル化したドメイン・オントロジー
- Resource Type (Ontology)
- 動機
 - Atlant-Hプロジェクトでは水素の国際市場の活動の自動分析ツールの開発を目指し、オントロジーベースの情報抽出とテキスト処理の根幹となる、水素経済の情報を構造化したオントロジーを開発。水素ドメインの知識を市場のインサイトが得られるようモデル化したい。
- リソースの設計方針
 - 水素市場の戦略的な予測のための検索、蓄積、配信の基盤となる知識ベース。
 - 対象概念は市場構造(市場のアクター、ロール、相互作用等)と分野の技術的知識を含む。
 - ソーステキスト中の語と概念のマッピングのため、語彙-意味レイヤーを含む。
- 新規性
 - 動的に変化する水素市場のインサイトを検索、蓄積、配信するための基盤としてのオントロジー
 - 既存のオントロジーは上記の知識を表現するには不十分。
- リソースの再利用性／活用実績
 - 上記プロジェクトでフランホーファーIISが水素市場のインサイトの検索、蓄積、配信に利用し、後続プロジェクトで使用予定。サードパーティの水素プロジェクトの利用も想定。
- 品質
 - LOTフレームワークに基づきオントロジーを開発。上記プロジェクトのユースケースとは別に、オントロジーのユースケースも定めている。テストケースによる機能要件への適合の評価等を実施。
- リソースの可用性
 - ソースコード、テストケース、要件定義、ユースケース等: <https://purl.org/holy/repository>
 - オントロジー、文書: <https://doi.org/10.5281/zenodo.7447958>
- 次にすべきこと
 - 上記プロジェクトでモデルの拡張、改良を計画(他の水素技術、バリューチェーンステージに拡張等)

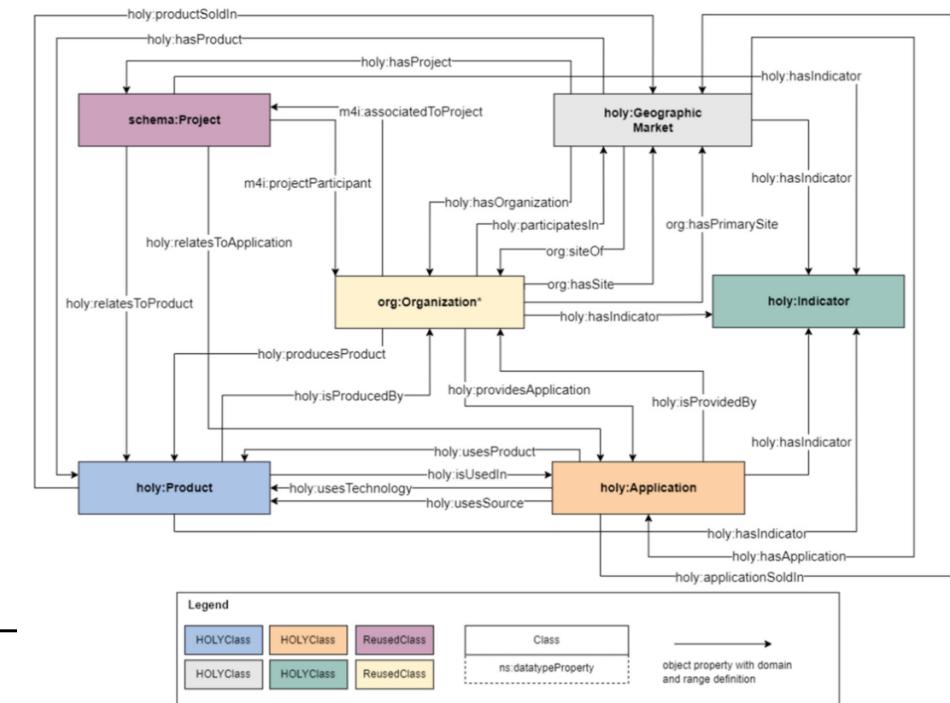


Fig. 3. Overview of the *Hydrogen Ontology*'s six main classes and related properties.

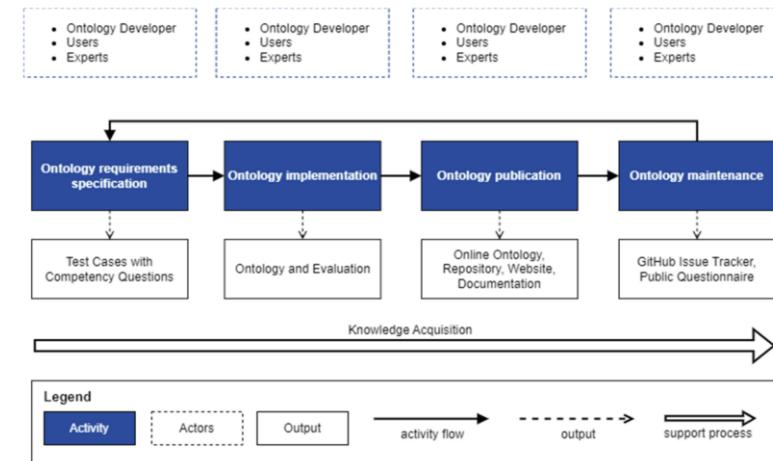


Fig. 1. The *Linked Open Terms (LOT)* Framework used to develop the *Hydrogen Ontology*.

AsdKB: A Chinese Knowledge Base for the Early Screening and Diagnosis of Autism Spectrum Disorder

Tianxing Wu^{1,2}, Xudong Cao¹, Yipeng Zhu¹, Feiyue Wu¹, Tianling Gong¹, Yuxiang Wang³, and Shenqi Jing^{4,5} /¹Southeast University, ²Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), ³Hangzhou Dianzi University, ⁴The First Affiliated Hospital of Nanjing Medical University, ⁵Nanjing Medical University

- 一言でいうと
ASD(自閉スペクトラム症)のスクリーニングと診断に必要な知識を含む中国の知識ベース
- Resource Type (Knowledge base, Ontology)
- 動機
ASDの診断確定にはスクリーニングと行動観察が必要だが、中国では専門医が不足し、早期に診断を受けられない子供が多い。スクリーニングと診断に必要な知識を含む知識ベースを開発。
- リソースの設計方針
ASDのスクリーニングと診断に必要なオントロジーと事実に関する知識を含む知識ベースを開発
 - 異種の情報源から疾患、診断、事実(専門医、病院等)に関する情報を収集(Fig.1参照)
 - オントロジーの主要な概念は人手で選定。既存の語彙を再利用。
- 新規性
ASDに関する初の中国の公開知識ベース
- リソースの再利用性／活用実績
AsdKBによる早期スクリーニングと診断のプロトタイプシステム(QA, スクリーニングスケールによるASDリスクの補助診断、専門医推薦)を実装
- 品質
 - accuracyの評価: $97.02\% \pm 1.21\%$ (サンプル: 約1%(732/69,290トリブル))
 - タスクによる評価: ヘルスケアサイトのASDに関するよくある質問(100問)の81%をカバー
- リソースの可用性
 - AsdKB: <https://w3id.org/asdkb/> (ontology: <https://w3id.org/asdkb/ontology/>)
 - データ: <https://zenodo.org/record/8199698>, 基本情報: <https://github.com/SilenceSnake/ASDKB>
- 次にすべきこと
 - データ品質向上(プロトタイプのログデータからの学習), 電子カルテの分析

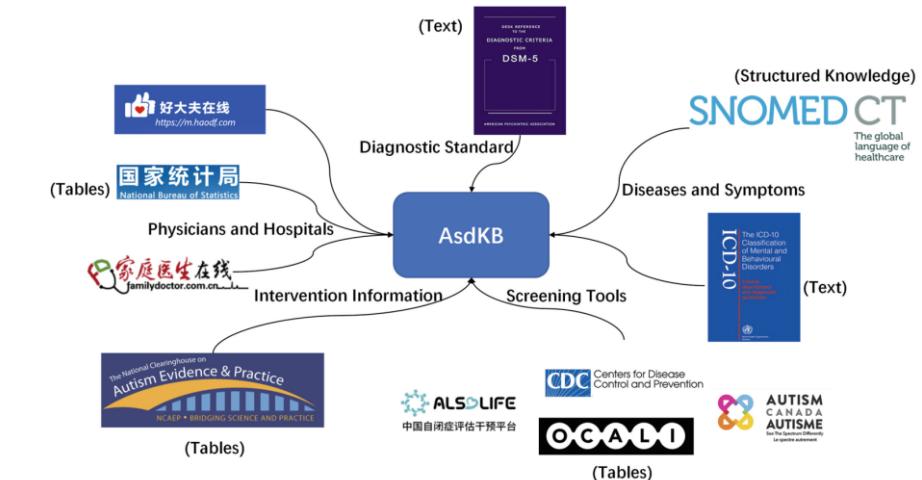


Fig. 1. The data sources for building AsdKB.

(a)

(b)

(c)

Fig. 7. An illustration of the auxiliary diagnosis system.

Benchmarking Geospatial Question Answering Engines Using the Dataset GEOQUESTIONS1089

Resource Track

Sergios-Anestis Kefalidis¹, Dharmen Punjani², Eleni Tsalapati¹, Konstantinos Plas¹, Mariangela Pollali¹, Michail Mitsios¹, Myrto Tsokanaridou¹, Manolis Koubarakis¹, and Pierre Maret^{2/1} National and Kapodistrian University of Athens, ² Université St. Monnet

- 一言でいうと
地理空間QAエンジンの評価用ベンチマークデータセットを作成, SOTAのエンジンを評価.
- Resource Type (Benchmark data)
- 動機
地理空間QAエンジンの有効性(effectiveness), 効率(efficiency)の評価のため, 意味的に複雑なもの等, 多様なタイプの質問を含む大規模なベンチマークデータセットを作成.
- リソースの設計方針
 - 対象: YAGO2, YAGO2geo(YAGO2のうち, 緯度経度情報をもつエンティティのサブセット)
 - 従来のデータセット(GEOQUESTIONS201)より複雑な質問を多く含む. 自然言語理解やGeoSPARQLの高度なfeatures(ネスト, no-existsフィルタ, 算術演算)の必要な質問等.
- 新規性
地理空間QAの最大のベンチマーク(1089トリプル), 意味的に複雑な質問を含む
- リソースの再利用性／活用実績
エンジン(GeoQA2, Hamzeiら(2021))を評価(複雑な質問には両者とも対応が不十分).
- 品質
クエリのタイプ別/作成方法別サブセットによりエンジンの詳細な分析が可能(報告者コメント).
- リソースの可用性 <https://github.com/AI-team-UoA/GeoQuestions1089>
- 次にすべきこと
地理空間QAエンジンの深層学習のトレーニングに使えるようデータセットを拡張

カテゴリ	説明 (3. The GEOQUESTIONS1089 Datasetの本文より引用, 太字は報告者による)
A	Asking for a thematic or a spatial attribute of a feature, e.g., " Where is Loch Goil located? "
B	Asking whether a feature is in a geospatial relation with another feature or features, e.g., " Is Liverpool east of Ireland? ".
C	Asking for features of a given class that are in a geospatial relation with another feature. E.g., " Which counties border county Lincolnshire? " or " Which hotels in Belfast are at most 2km from George Best Belfast City Airport? "
D	Asking for features of a given class that are in a geospatial relation with any features of another class, e.g., " Which churches are near castles? ".
E	Asking for features of a given class that are in a geospatial relation with an unspecified feature of another class, and either one or both, is/are in another geospatial relation with a feature specified explicitly. E.g., " Which churches are near a castle in Scotland? " or " In Greece, which beaches are near villages? "
F	As in categories C, D and E above, plus more thematic and/or geospatial characteristics of the features expected as answers, e.g., " Which mountains in Scotland have height more than 1000 m? "
G	Questions with quantities and aggregates, e.g., " What is the total area of lakes in Monaghan? " or " How many lakes are there in Monaghan? ".
H	Questions with superlatives or comparatives, e.g., " Which is the largest island in Greece? " or " Is the largest island in France larger than Crete? ".
I	Questions with quantities, aggregates, and superlatives/comparatives, e.g., " Which city in the UK has the most hospitals? " or " Is the total size of lakes in Greece larger than lake Loch Lomond in Scotland? "

1A: Ontologies and Knowledge Graphs I

7A: Ontologies and Knowledge Graphs III

東京都立大学 太田葵

AIDA-Bot 2.0: Enhancing Conversational Agents with Knowledge Graphs for Analysing the Research Landscape

Antonello Meloni, Simone Angioni, Angelo Salatino, Francesco Osborne, Aliaksandr Birukou, Diego Reforgiato Recupero, and Enrico Motta (University of Cagliari, The Open University, University of Milano Bicocca, Springer-Verlag GmbH)

- 一言でいうと

論文データ等のKGを活用した科学技術文献専用チャットボット

- 動機

多くの人に責任が分散している状態で科学技術文献の分析・サーバイをするのは困難

- 手法

文献情報の大規模なKGを活用し研究関係の幅広い質間に答えるチャットボットの作成

- 背景

2021年に発表したAIDA-Botの改良版

- 結果

事前に設定された質問とopen question 両方に対応できるボットが作成できた

- 利用実績

おそらくuser studyのみ

- 学んだ教訓やベストプラクティス

GPTの登場によりチャットボットに求められるレベルそのものがあがってきた

Aviation Certification Powered by the Semantic Web Stack

Paul Cuddihy, Daniel Russell, Eric Mertens, Kit Siu, Dave Archer, and Jenny Williams

- 一言でいうと

軍事研究プログラムの下で、W3Cセマンティック技術を応用した効率的な証拠のキュレーションが可能であることを実証

- 動機

ソフトウェアシステムが大きくなりすぎて認証と証拠の量が追いつかない

- 手法

W3CセマンティックWeb技術の適用：異種データの統合

- 背景

すべての航法システムは各種規定に沿っているかどうかの審査を通過する必要がある

- 結果

グラフビューと再帰的グラフクエリが目的を果たしている

- 利用実績

記載見当たらず

- 学んだ教訓やベストプラクティス

ユーザが詰まるのはトリプル作成の部分

Scaling Data Science Solutions with Semantics and Machine Learning: Bosch Case

Baifan Zhou, Nikolay Nikolov, Zhuoxun Zheng, Xianghui Luo, Ognjen Savkovic, Dumitru Roman, Ahmet Soylu, and Evgeny Kharlamov

- 一言でいうと

Sem-Cloudを提案：ユーザ，ETL，データ分析，クラウドインフラを仲介する意味的なまとめ

- 動機

スマート工場化により熟練者以外のクラウドの需要が高まってきた一方、彼らをトレーニングするのは時間が無いため難しい

- 手法

分散型クラウドコンピューティングでのセマンティックデータ統合とデータ分析を提案し、非熟練者のデプロイを可能にした

- 背景

Boschの1工場では1か月あたり1.9 millionを超える接続記録が様々なソフトウェア環境から発生、クラウド処理

のニーズ高まる

- 結果

クラウド使用経験がなかった人たちが、Sem-Cloudの使用によりクラウドに関する知識を深めることができた

- 利用実績

Boschの様々な職種による上記のテストのみ

- 学んだ教訓やベストプラクティス

開発初期段階でコストが最も多くかかる

Solving the IoT Cascading Failure Dilemma Using a Semantic Multi-agent System

Amal Guittoum ,Francois A "issaoui ,Sebastien Bolle , Fabienne Boyer , and Noel De Palma

- 一言でいうと
Cooperative Multi-agent System をベースとした、相互独立したIoTデバイス群の異常を自動的かつ協調して解決するアプローチを提案
- 動機
相互独立したIoTデバイスを管理（Device Management, DM）するのは、バグが複数個所にわたるため困難
- 手法
DMプラットフォームと統合して利用できるSemantic AgentのOSAMAを適用
デバイスの相互依存関係を共有できるオントロジーを含有
- 背景
 - 各IoTデバイスでそれぞれ異なる管理体制・メンテナンス方法
 - IoTデバイスどうしの依存関係がバグを引き起こす
- 結果
質的評価：異常原因究明までに5秒（先行デバイスでは20秒ほど）
- 利用実績
評価実験（スマートホームでの実験）のみ
- 学んだ教訓やベストプラクティス
修理にかかる時間やコストのロスを可能な限り抑えられる

TEC: Transparent Emissions Calculation Toolkit

Milan Markovic , Daniel Garijo , Stefano Germano, and Iman Naja

- 一言でいうと

温室効果ガス排出量を求めることができるオープンソースツールキット

- 動機

それぞれの活動によって排出される温室効果ガスの量を求めるのは困難

- 手法

排出量と排出源を示す2種類のオントロジー, 変換ファクターKG, プロトタイプ

- 背景

世界40か国以上が（自国の）組織に対し定期的な温室効果ガス排出量の測定と報告を求めている

- 結果

商用のnon-semanticなデータ収集方法の代替として機能し, 将来的なソフトウェア開発の道を提示

- 利用実績

記載なし

- 学んだ教訓やベストプラクティス

記載なし

The SAREF Pipeline and Portal—An Ontology Verification Framework

Maxime Lefrancois and David Gnabasik

- 一言でいうと

Smart Applications REference Ontology の開発パイプラインとワークフローの説明

- 動機

ETSI STFが“ユーザエンゲージメントのためのSAREF開発フレームワーク・ワークフローとコミュニティポータルの開発”を完了

- 手法

- すべてのSAREFプロジェクトが遵守すべきガイドラインを自動的にチェック
- 各プロジェクトのドキュメントを提供
- SAREF全体をGitで管理する構造を提供

- 背景

SAREF：様々なIoTソリューション間の意味的な相互運用を可能にするオントロジー群

- 結果

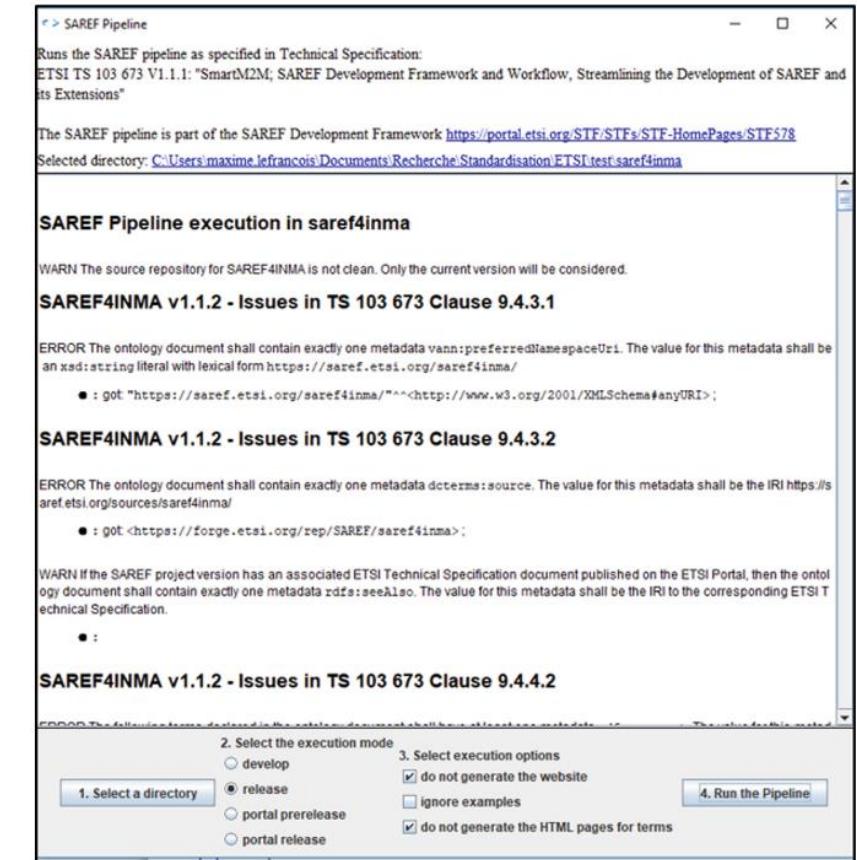
各SAREFオントロジーの質の向上に貢献

- 利用実績

すでにユーザによって利用されており、上記の結果

- 学んだ教訓やベストプラクティス

各名前空間の意味を覚えるのはユーザにとってたびたび困難



Industry: Ontologies and Knowledge Graphs II

10B: Linked Data in action

情報・システム研究機構 山本泰智

Improving Reasoning on Large Ontologies via Ontology Modularity

Jieying Chen (1,2), Johanna P. Haarseth (1), Christian M. Hansen (3), Martin G. Skjæveland (1), Arild Waaler

1 Department of Informatics, University of Oslo, Oslo, Norway

2 Department of Computer Science, University of Oxford, Oxford, UK

3 Aibel AS, Asker, Norway

- **一言でいう**

石油関連施設を建設するには多くの関係者が必要であり、そこでやりとりされる書類の量も膨大なので、その作業の効率性と交わされるデータの正確性を高めるために、AibelはMaterial Master Data (MMD) オントロジーを構築した。

- **動機**

石油関連施設建設に伴う多数関係者間でのデータの融通が非効率であった。

- **課題**

データと情報のやりとりを円滑にするために多くの書類が作られてやりとりされる。

- **手法**

OWL2に基づくMaterial Master Data (MMD) オントロジーを構築し、頻繁に起こる規制の改正に対応しやすいように、OWL2推論器で全体の一貫性を検証可能。さらに、全体のオントロジーに対する問合せでは時間が30分近くかかるので、用途ごとに必要なクラスとプロパティを見極めた部分オントロジーを構築した。

- **評価**

部分オントロジー（モジュール）を用いることで、各用途における問合せが長くても14秒程度までに効率化できた。

- **考察**

部分オントロジーを用いても、オントロジー全体に対して問い合わせを行った場合と同一の結果が得られるようにしているため、MMDオントロジー全体で対象領域を記述するとともに、各用途に応じた部分オントロジーを用いることで、実用性を担保できた。

- **知見**

今後は、さらなる用途を見つけ、セマンティックウェブ技術が適用できる範囲を広げていきたい。

A Knowledge Graph-based Approach for the Quality Management of Bosch Products

Qiushi Cao (1), Irlán Grangel-González (2), Lin Du (1)

1 Corporate Research, Bosch (China) Investment Ltd., Shanghai, China

2 Corporate Research, Robert Bosch GmbH, Renningen, Germany

- **一言でいうと**

Boschの電化製品に発生する内部欠陥を解析するためのデータが多様でバラバラなので、知識グラフを構築して全てのデータの透明性を高め、解釈しやすい意味を持たせ、アクセスしやすくした。

- **動機**

欠陥に関連するデータの多様性と、データのサイロ化。

- **課題**

これまでのデータ管理への取り組みでは欠陥に繋がるデータの効率的な解析が行えなかった。

- **手法**

関連データの解析に至るまでの処理をデータソース層からアプリケーション層までの4つの層に整理する。そして、データソース層に当たる対象データ群からオントロジーにマップして知識グラフを構築することでデータを統合し、解析などのアプリケーション層で利用しやすい状態にする。

- **評価**

知識グラフを用いたデータ統合を行うことで、これまで3ヶ月かかっていた内部欠陥を見つける処理が、3分で済むようになった。さらに、知識グラフを用いた手法により、データ解析効率が70%改善し、結果としてボッシュの工場における大幅な費用削減につながった。金額に換算すると、年当たりで20万ユーロになる。

- **考察**

提案手法は、広く知識グラフを用いた取り組みに適用できる。

- **知見**

知識グラフはデータベース間の意味的衝突（SICs : Semantic Interoperability Conflicts）を低減するのに役立つ。今後はアプリケーション層として、内部欠陥予知機能を開発したい。

Aerospace Qualification Services Knowledge Graph: A Leap towards Enhanced Data Management

Raed Awill (1,2), Wajahat Ali Khan (1), Maqbool Hussain (1), Muhammad Sadiq Hassan Zada (1,2), Ben Anderson (2)

(1) University of Derby, Derby, UK

(2) AddQual LTD, Derby, UK

- **一言でいうと**

航空宇宙業界では大量のデータが必要になり、これまで利用していた関係データベースでは限界があったので、知識グラフ化した。

- **動機**

航空宇宙業界で利用するデータは膨大で、これまで関係データベースを利用していたが、データの一貫性の維持しながらスキーマ変更を行うことが限界に達した。

- **課題**

関係データベースでは解決できない程にデータが複雑であった。

- **手法**

データ取得から始め、知識モデリングと表現、データの永続化、そしてデータ可視化の順に処理を進める。

- **評価**

構築した知識グラフAQS-KGは関係データベースの問題を克服し、水平型の開発と相互運用性を促進する。

また、同一の意味である100の名前を持つ特徴が特定された。

- **考察**

AddQualにより構築されたAQS-KGは、クエリ応答時間を最大35%減らせた。

- **知見**

パーティの数が2022年の1000から2026年の20000に増加が見込まれているため、クエリ処理時間の効率化は必須。

Link Traversal Query Processing Over Decentralized Environments with Structural Assumptions

Ruben Taelman, Ruben Verborgh/Ghent University, Belgium

- 一言でいうと
大規模に散在する個人情報データベースに対して効率よく所望のデータを取得するためのクエリ処理方法LTQPの拡張機能の提案

- 動機

Solidが普及するためには多くの、比較的小規模で非常に多くのRDFデータセットへの効率的な問い合わせ手法が必要。

- 手法

LTQP(Link Traversal Query Processing)という、特定のデータセットから、そこに含まれるリンクを次々に辿って必要なデータを取得する手法を拡張して効率を向上。

- 背景

Solidでは個人情報が個々人の手元に置かれるので、予め索引を作つておくことはできないから、動的に効率よく必要なデータを取得する手法の開発が必須。

- 結果

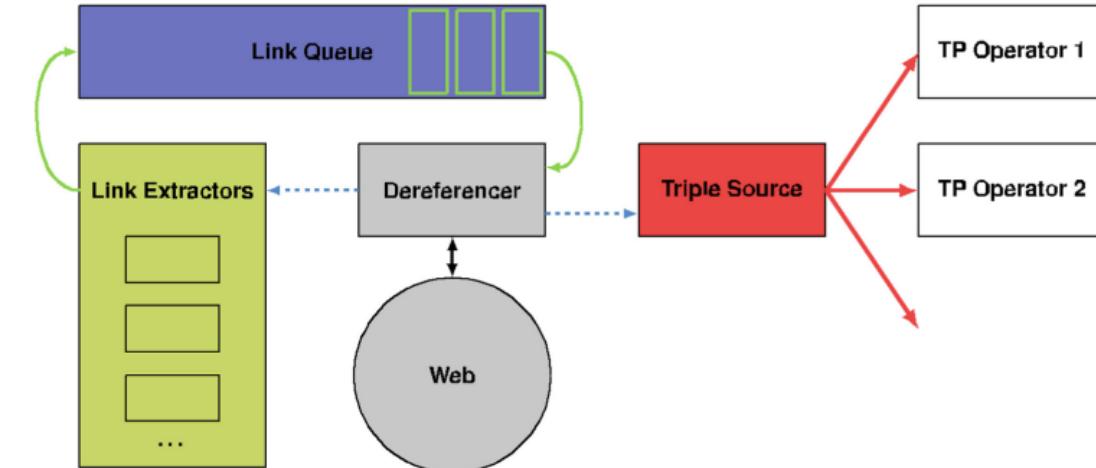
RDFのタイプ情報 (rdf:type) など、予め必要なデータが含まれるデータの形式を用意しておき、それに基づき効率的にデータを収集できた。

- 考察

今回の研究では、多くのクエリが一秒未満で結果が得られた。

- 課題

複雑なデータを効率よく取得するには、更なるクエリ計画手法の改善が必要。



LDkit: Linked Data Object Graph Mapping Toolkit for Web Applications

Karel Klíma (1), Ruben Taelman (2), Martin Nečaský (1) / 1. Charles University, Prague, Czechia, 2. Ghent University –IMEC, Ghent, Belgium

- 一言でいうと
LDkitと呼ばれる、Linked DataをTypeScriptに馴染んだ形式に変換するObject Graph Mapping (OGM) フレームワークの提案。
- Resource Type
Software Framework ? (記載なし)
- 動機
多くのウェブ開発者に馴染みやすい形式でLinked Dataを利用可能にしたい。
- リソースの設計方針
RDFデータモデルを抽象化して、Linked Dataをウェブアプリケーションで直接利用することを容易にするLDkitの開発。
- 新規性
Linked Dataをウェブアプリケーションで簡単かつ直感的に利用可能な開発者向けのツールを提供。
- リソースの再利用性／活用実績
チコ政府で利用されている（<https://slovnik.gov.cz/>）など
- 品質
実用できな速度で処理されることを確認。
- リソースの可用性
<https://github.com/karelklima/ldkit>
- 次にすべきこと
RDFデータからLDkitスキーマや全体のフロントエンドアプリケーションを支援するためのツールの提供。

```
import { createLens } from "ldkit";
import { dbo, xsd } from "ldkit/namespaces";

const PersonSchema = {
  "@type": dbo.Person,
  "name": dbo.birthName,
  "birthDate": {
    "@id": dbo.birthDate,
    "@type": xsd.date,
  },
} as const;

const Persons = createLens(PersonSchema, {
  sources: ["https://dbpedia.org/sparql"]
});

const adaIri = "http://dbpedia.org/resource/Ada_Lovelace";
const ada = await Persons.findByIri(adaIri);

console.log(ada.name); // "The Hon. Augusta Ada Byron"
console.log(ada.birthDate); // Date object of 1815-12-10
```

Linked Data Objects (LDO): A TypeScript-enabled RDF Devtool

Jackson Morgan/O.team

- 一言でいうと
Linked Data Objects (LDO) と呼ばれるJavaScript (TypeScript) ベースでRDFデータの読み書きできる開発ツールを紹介。

- Resource Type**
Software Framework

- 動機**
RDFデータを処理するプログラムを開発するのにウェブ開発者に馴染みやすい枠組みが必要。

- リソースの設計方針**
RDFデータの読み書きに関するプログラムを、TypeScriptで開発しやすいように。

- 新規性**
TypeScriptを用いたRDF処理プログラム開発環境は新規。

- リソースの再利用性／活用実績、リソースの可用性**
License: MIT License
Permanent URL: <https://purl.archive.org/o.team/lld>
Canonical Citation: <https://doi.org/10.5281/zenodo.7909200>
17 stars on GitHub and 1,719 total downloads from NPM (May 8th, 2023)

- 品質**
ShExに基づく型指定やデータ処理などを、多くのウェブ開発者が使う枠組みで提供。
- 次にすべきこと**
JSON-LD contextへの対応やShEx検証機能の追加など

```

import React, { FunctionComponent } from 'react';
import { Header } from './Header';
import { Blog } from './Blog';

export const App: FunctionComponent = () => {
  return (
    <div className="App">
      <Header />
      <Blog />
    </div>
  );
}

PREFIX ex: <https://example.com/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ns: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
ex:FoafProfile EXTRA a {
  foaf:name xsd:string
  // rdfs:comment "A profile has 1 names" ;
  foaf:title ns:langString *
  // rdfs:comment "A profile has 0~∞ titles, and they could have translations" ;
  foaf:knows @ex:FoafProfile *
  // rdfs:comment "A profile has 0~∞ friends." ;
}

```

4A: Link Prediction I

市瀬 龍太郎

Comprehensive Analysis of Freebase and Dataset Creation for Robust Evaluation of Knowledge Graph Link Prediction Models

Nasim Shirvani Mahdavi, Farahnaz Akrami, Mohammed Samiul Saeef, Xiao Shi, Chengkai Li / University of Texas at Arlington

- 一言でいうと
Freebaseを用いた新たなデータセットの提案とその評価
- Resource Type**
データセット
- 動機**
既存のFreebaseデータセットでは、Freebaseの3つの特性を考慮していない
- リソースの設計方針**
Freebaseには、逆関係の付与、多項関係ノード（CVT）の導入、強力な型情報という他の知識グラフにない特性があるため、その影響を排除
- 新規性**
リンク予測研究における新たな評価手法の確立
- リソースの再利用性／活用実績**
新たなデータセットを5つのリンク予測手法で評価
- 品質**
既存のデータセットの問題点を丁寧に分析して、新たなデータセットを作成
- リソースの可用性**
<https://github.com/idirlab/freebases>
- 次にすべきこと**
リンク予測研究における新たな標準データセットとしての利用

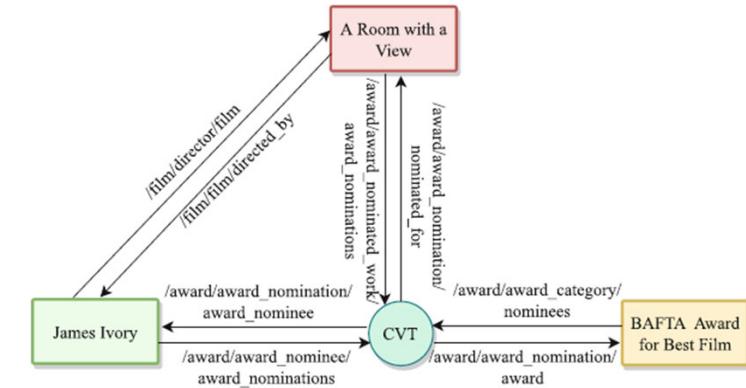


Fig. 1. A small fragment of Freebase, with a mediator node

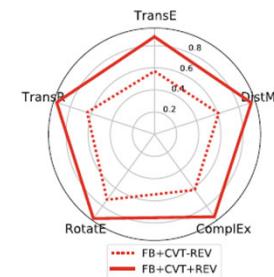


Fig. 2. Impact of reverse triples on MRR^{\uparrow} of embedding models

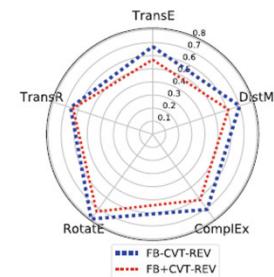


Fig. 3. Impact of mediator nodes on MRR^{\uparrow} of embedding models

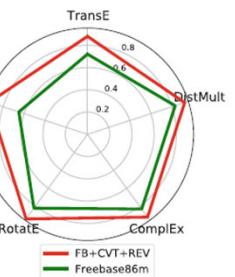


Fig. 4. Impact of non-subject matter triples on MRR^{\uparrow} of embedding models

Causal Inference-based Debiasing Framework for Knowledge Graph Completion

Lin Ren, Yongbin Liu, Chunping Ouyang / University of South China

- 一言でいうと
知識グラフ補完問題において、精度高く補完する手法の提案
- 動機
既存のリンク予測手法では、in-depthバイアス、in-breadthバイアスという2つのバイアスにより、不正確な結果を出力
- 手法
テキスト情報、知識グラフを利用して、因果分析を行うことで、適切なリンク予測候補を選出
- 背景
欠落したトリプルを精度高く補完し、知識グラフの構築方法を強化
- 結果
4つのデータセットで検証した結果、Hit@1で1～3%程度改善
- 考察
バイアスがみられるデータセットでは、より大きな改善
- 課題
知識グラフ補完問題の包括的なバイアスの分析と手法の一般化

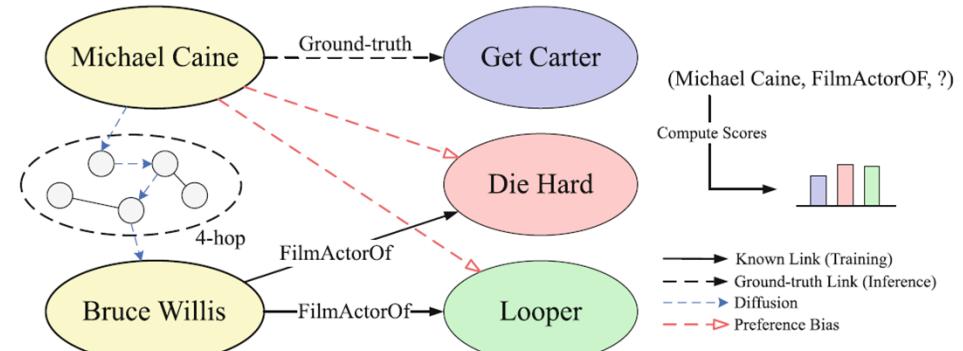
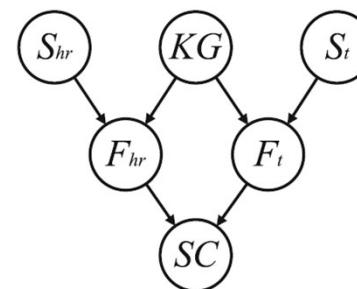


Fig. 1. A simple example from dataset FB15k-237 [35] for in-depth bias. *Diffusion* means the correlation diffusion along known links.



KG: Knowledge Graph Structure
 $S_{hr}(S_t)$: Semantic Information of $hr(t)$
 $F_{hr}(F_t)$: Feature Representation of $hr(t)$
SC: Score of t for hr

Fig. 4. The causal graph of knowledge graph completion. KG can be seen as the training set; S_{hr} and S_t are the text description of corresponding entities and relations. It should be noted that unobserved biases which can not be observed and interfered with are ignored in this figure.

CapsKG: Enabling Continual Knowledge Integration in Language Models for Automatic Knowledge Graph Completion

Janna Omeliyanenko, Daniel Schlör, Albin Zehe, Andreas Hotho / Julius-Maximilians-University Würzburg

- 一言でいうと

言語モデルを用いた知識グラフ補完手法に対して、継続的に予測モデルの学習を可能とする手法の提案

- 動機

異なる関係のリンク予測に対して、言語モデルのファインチューニングを繰り返すと、以前学習した関係のリンク予測が精度が下落

- 手法

関係毎に言語モデルの隠れ層の情報を保持しておくと同時に、その情報を変更できるような仕組みを導入

- 背景

新しい関係の導入など、知識グラフ補完のために継続的にモデルを学習する必要性

- 結果

表3 : CapsKG > BERT > BERT-CL

- 考察

継続的に学習をしても性能を維持し、学習したことの壊滅的な忘却を回避可能

- 課題

様々な言語モデルに対して系統的に評価

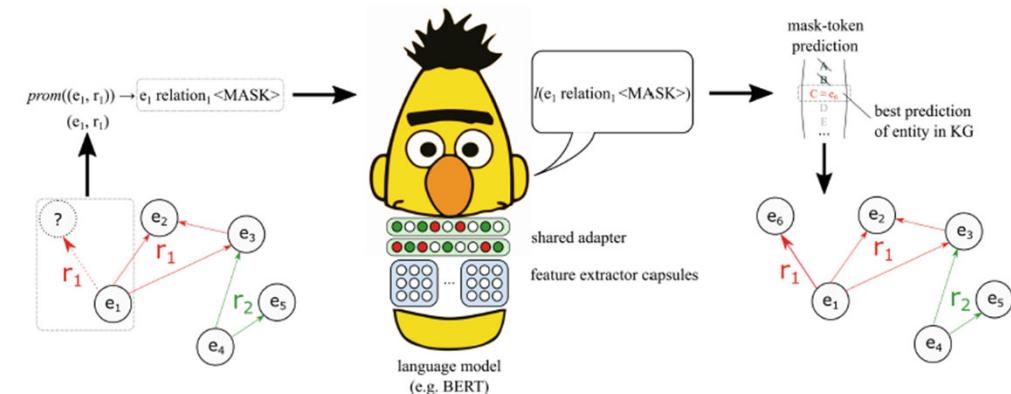


Table 3. Results for link prediction on WN18, YAGO, and FB15k with 5 random seeds, reporting mean and standard deviation of Hits@1 performance.

Model/Dataset	WN18	YAGO	FB15k
BERT _{frozen}	10.7 ± 0.0	27.0 ± 0.0	5.5 ± 0.0
BERT	25.8 ± 0.5	48.0 ± 0.1	34.7 ± 0.7
Adapter	26.4 ± 0.5	48.7 ± 0.4	35.6 ± 0.2
BERT-CL	20.2 ± 1.2	40.4 ± 4.0	11.4 ± 3.2
Adapter-CL	18.4 ± 1.2	41.5 ± 0.7	16.3 ± 1.4
CapsKG _{forward}	27.4 ± 0.4	49.4 ± 0.3	36.1 ± 0.3
CapsKG _{backward}	27.2 ± 0.4	49.5 ± 0.3	34.9 ± 0.3

6C: Link Prediction II

市瀬 龍太郎

Spatial Link Prediction with Spatial and Semantic Embeddings

Genivika Mann, Alishiba Dsouza, Ran Yu, Elena Demidova / University of Bonn, Lamarr Institute for Machine Learning and Artificial Intelligence

- 一言でいうと
地理情報を対象としたリンク予測問題を解決する手法の提案
- 動機
グラフ構造からリンク予測を行うと、地理的な空間情報が反映されない
- 手法
位置情報、ラベル情報などの埋め込みを作成し、そこからリンク予測を実施する
2つの手法を提案
- 背景
地理知識グラフWorldKGでは、リンクがされていない多くのエンティティが存在
- 結果
従来のリンク予測手法よりも大幅に性能が改善
- 考察
トリプルに加え、空間情報、文字情報の利用により、性能が改善
- 課題
地理に関する複雑な質疑応答に対する、空間情報と文字情報の埋め込みの応用

10021976	
Download triples	
Property	Object
rdfs:label	Leicester
rdf:type	wkgs:City
wkgs:isIn	Leicestershire, England, UK
wkgs:isInContinent	Europe
wkgs:isInCountry	<u>United Kingdom</u>
wkgs:isInCounty	Leicestershire
wkgs:nameAr	لیکسٹر
wkgs:nameCy	Caerlŷr
wkgs:nameJa	レスター
wkgs:nameKn	ලේස්ටර්
wkgs:nameLt	Lesteris
wkgs:nameRu	Лестер

838090640	
Download triples	
Property	Object
rdfs:label	United Kingdom
rdf:type	wkgs:Country
wkgs:officialNamePl	Zjednoczone Królestwo Wielkiej Brytanii i Irlandii Północnej
wkgs:nameAr	المملكة المتحدة
wkgs:nameBe	Великобритания

302324104	
Download triples	
Property	Object
rdfs:label	Leicestershire
rdf:type	wkgs:County
wkgs:isInCountry	<u>UK</u>
wkgs:osmLink	https://www.openstreetmap.org/node/302324104
wkgs:source	GNS

Table 2. Spatial link prediction performance in transductive setting. The best results are highlighted in bold, and runner-up results are underlined.

Model	TD1 Dataset						TD2 Dataset					
	MR	MRR	H@1	H@3	H@5	H@10	MR	MRR	H@1	H@3	H@5	H@10
TRANSE	1104	0.147	0.065	0.174	0.229	0.306	12594	0.199	0.114	0.237	0.286	0.357
DISTMULT	1328	0.861	<u>0.780</u>	0.934	0.968	0.977	22005	0.660	0.568	0.747	0.799	0.808
HAKE	1592	0.327	0.214	0.426	0.475	0.499	13194	0.305	0.235	0.366	0.402	0.409
ROTATE	1730	0.791	0.743	0.831	0.849	0.865	31671	0.010	0.002	0.004	0.006	0.021
COMPLEX	34410	0.005	0.002	0.005	0.007	0.012	41360	0.025	0.016	0.026	0.032	0.041
COMPGCN	750	0.095	0.000	0.004	0.078	0.396	17508	0.229	0.020	0.396	0.485	0.580
CONVKB	2555	0.099	0.019	0.132	0.179	0.249	11348	0.247	0.022	0.405	0.531	0.655
RESCAL	446	0.006	0.000	0.001	0.001	0.000	12337	0.002	0.000	0.000	0.000	0.000
LITERALE	70129	0.000	0.000	0.000	0.000	0.000	43276	0.022	0.016	0.018	0.022	0.039
L2E-DISTMULT	986	<u>0.936</u>	<u>0.919</u>	0.951	0.958	0.963	7113	0.797	0.776	0.811	0.820	0.830
SSLP	144	0.894	0.823	<u>0.962</u>	<u>0.972</u>	<u>0.978</u>	6505	<u>0.819</u>	<u>0.805</u>	<u>0.831</u>	<u>0.836</u>	0.844
USLP	188	0.964	0.941	0.990	0.995	0.996	68	0.856	0.845	0.862	0.866	0.872

Textual Entailment for Effective Triple Validation in Object Prediction

Andrés García-Silva, Cristian Camilo Berrío Aroca, Jose Manuel Gomez-Perez / Expert.ai

- 一言でいうと
テキストからトリプルを抽出して知識グラフを拡張する手法の提案
- 動機
言語モデルによりトリプルを作成すると意図しない結果やハリシネーションを起こす可能性
- 手法
言語モデルで作成したトリプルをWeb検索した文と比較して含意関係認識技術により妥当性を検証して候補を決定
- 背景
言語モデルの発達と、含意関係認識技術の向上
- 結果
ベースライン手法よりも、含意関係認識を利用した方が性能が向上
- 考察
既存の知識グラフとテキストメッセージのNERを候補オブジェクトのソースとして使用し、トリプルの検証を行った場合に、全体として最高の性能を達成
- 課題
より大きなパラメータを持つ言語モデルにおける効果の検証、含意関係認識モデルの改善

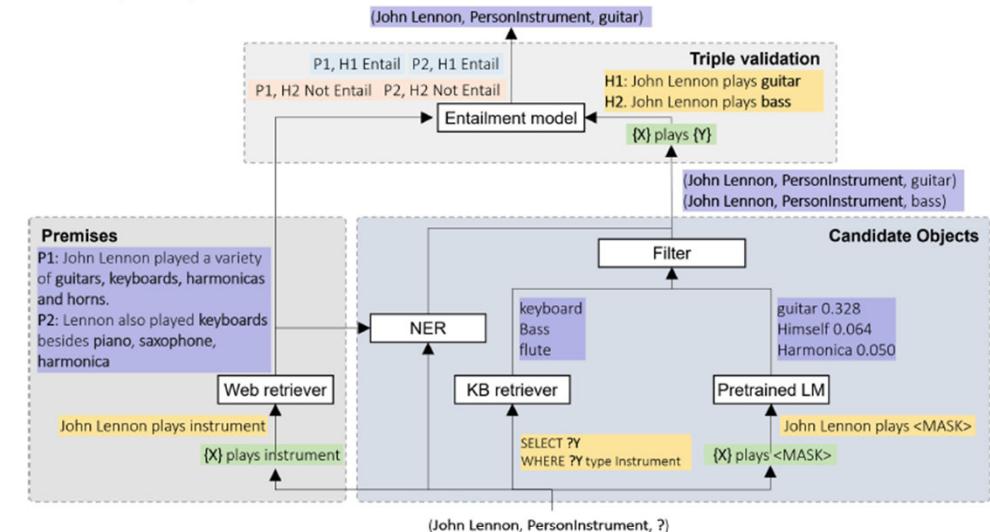


Fig. 1. Table 1. Object prediction using pretrained models vs. such models additionally trained in different scenarios. The models include the LM from where we get objects and the entailment model that we use for triple validation. The SATORI versions evaluated include three sources of candidate objects (Language Model LM, Knowledge Graph KG, Named Entities NER) and three entailment models (DeBERTa xs, BERT large, DeBERTa xl).

SATORI		Pretrained			Training Data											
					5%			10%			20%			100%		
RTE Model	Obj. Source	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
DeBERTa xs	LM	83.6	48.6	47.9	80.5	51.7	47.9	81.9	51.5	48.4	84.1	51.3	50.3	91.3	44.7	46.8
BERT large	LM	82.1	49.9	48.6	79.9	52.2	48.1	81.7	51.8	48.7	84.8	51.2	50.6	89.3	46.0	47.1
DeBERTa xl	LM	83.0	49.5	48.5	82.4	51.9	48.5	84.3	52.1	50.1	87.0	52.0	51.4	91.1	47.8	48.4
DeBERTa xs	KG+NER	67.7	61.8	55.2	67.0	60.9	52.1	69.5	60.1	52.6	71.0	60.8	53.8	73.5	60.4	54.7
BERT large	KG+NER	67.0	62.9	55.2	68.2	61.8	52.2	70.0	61.1	52.9	69.9	63.0	54.9	66.9	64.8	55.1
DeBERTa xl	KG+NER	70.0	62.5	55.9	71.4	62.2	54.9	72.6	62.3	55.1	70.7	63.6	56.5	68.3	64.0	54.8
DeBERTa xs	LM+KG+NER	66.6	62.0	54.4	65.4	61.2	51.5	67.1	60.8	51.8	69.1	61.5	53.4	73.3	60.5	54.6
BERT large	LM+KG+NER	64.5	63.2	54.7	65.7	62.6	51.5	68.2	61.6	52.7	68.6	63.5	54.8	67.2	65.1	55.7
DeBERTa xl	LM+KG+NER	63.5	64.3	53.9	69.4	62.6	54.2	70.6	63.1	54.8	69.5	64.2	56.5	70.0	64.3	56.6
LM-baseline		68.1	46.7	43.2	59.9	50.8	43.3	63.6	49.0	44.3	67.2	49.6	46.9	52.2	48.1	41.4
QA-baseline		67.1	42.3	40.9	54.1	45.2	38.5	57.9	45.4	40.6	57.6	46.5	41.7	57.4	52.7	47.3
RE-baseline		85.5	33.3	31.8	83.8	41.8	40.6	81.3	43.0	42.0	78.6	44.7	43.2	70.0	54.9	48.2

Literal-Aware Knowledge Graph Embedding for Welding Quality Monitoring: A Bosch Case

Zhipeng Tan, Baifan Zhou, Zhuoxun Zheng, Ognjen Savkovic, Ziqi Huang, Irlan Granel Gonzalez, Ahmet Soylu, Evgeny Kharlamov / Bosch Center for AI, RWTH Aachen University, University of Oslo, Oslo Metropolitan University, Free University of Bozen-Bolzano

- 一言でいう
溶接機器のモニタリングに対して知識グラフの手法を適用した報告
- 動機
溶接時に生成される大量のデータから、品質管理に必要な溶接スポットの直径、溶接個所がどの車体の部分かを知りたい
- 手法
データを知識グラフに変換し、埋め込みを用いたリンク予測問題として定式化
- 背景
製造業において知識グラフ埋め込みがどの程度まで利用できるかを評価する試み
- 結果
MLP, RotatE, AttHよりもTransEを使う方が良い結果
- 利用実績
ボッシュの取り組み
- 学んだ教訓やベストプラクティス
溶接直径予測はMLPよりもTransEを使った方がよいが、産業用アプリにはまだ不十分。車体予測は完全ではないが有望

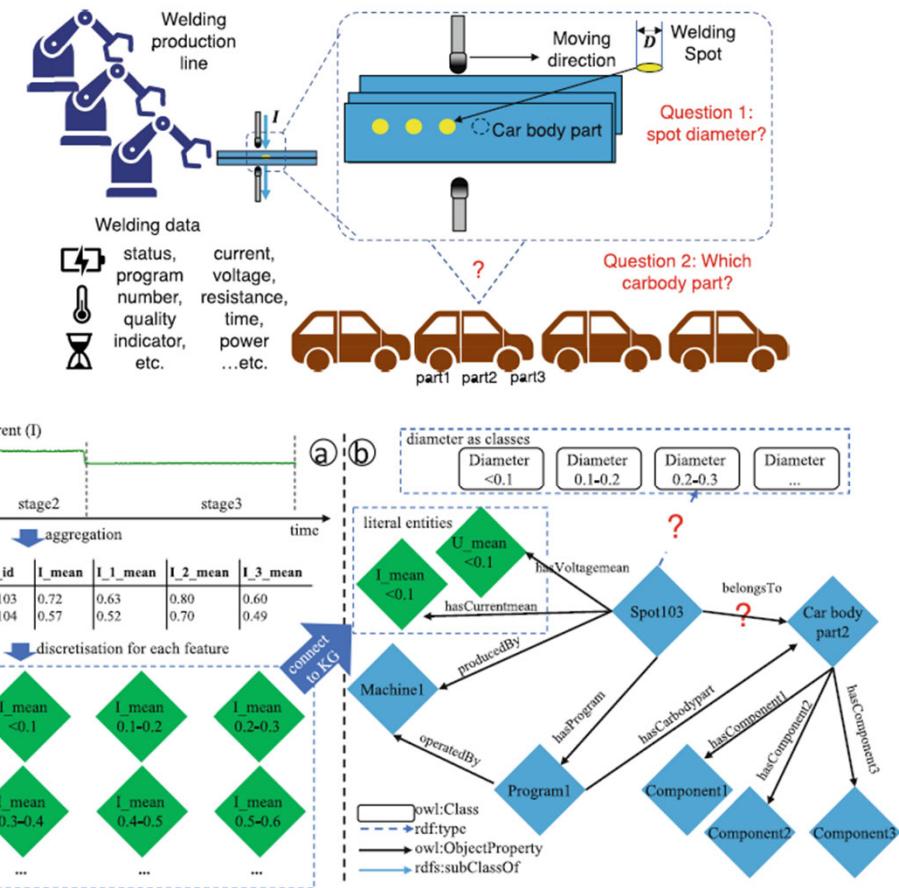


Fig. 4. (a) Procedure of literal embedding (b) Partial illustration of the welding KG

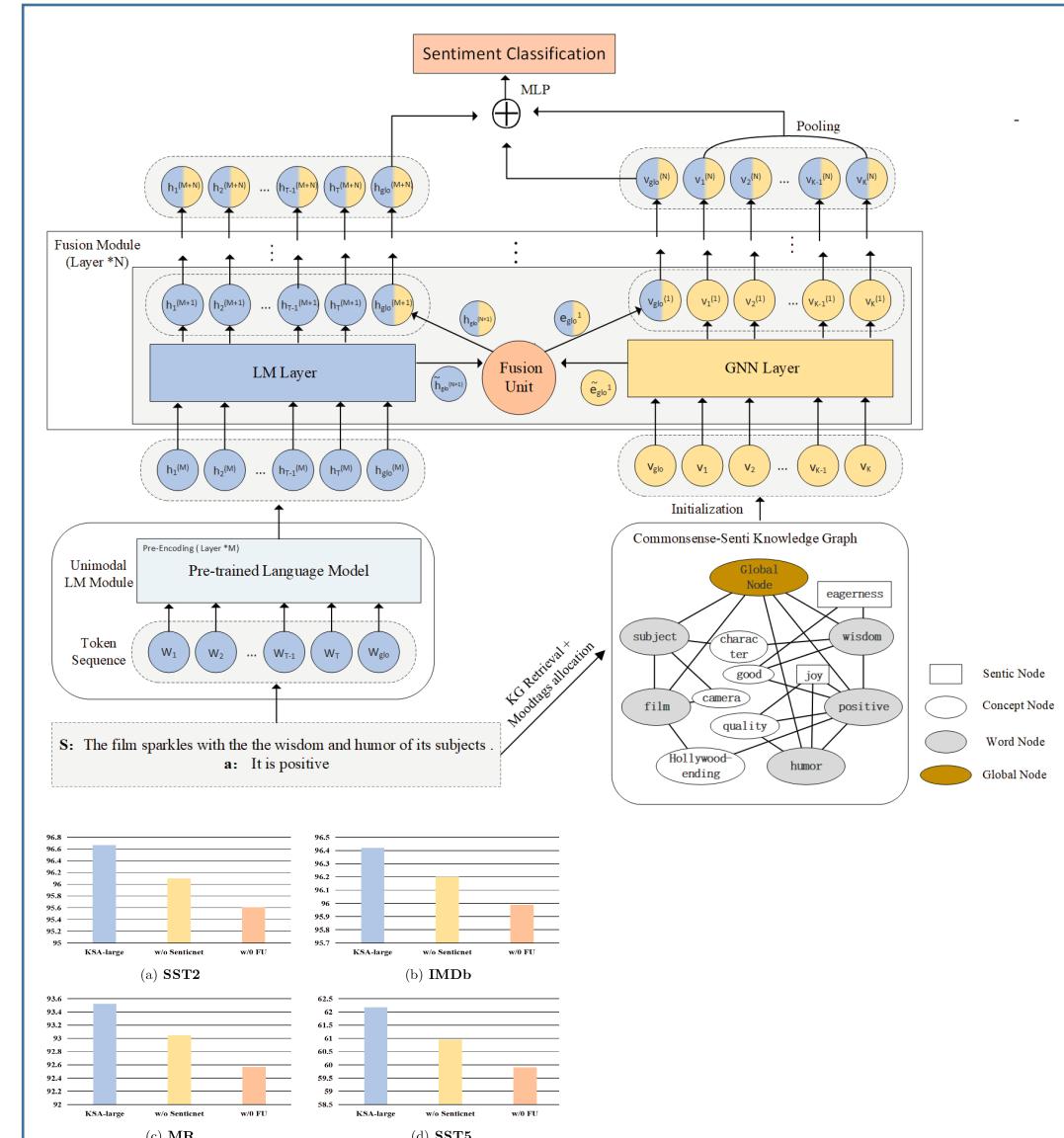
Session 3A: Knowledge Graph Embeddings I

吉丸 直希

Knowledge Graph Enhanced Language Models for Sentiment Analysis

Jie Li, Xuan Li, Linmei Hu, Yirui Zhang and Jinrui Wang/Beijing University of Posts and Telecommunications, Beijing, China.

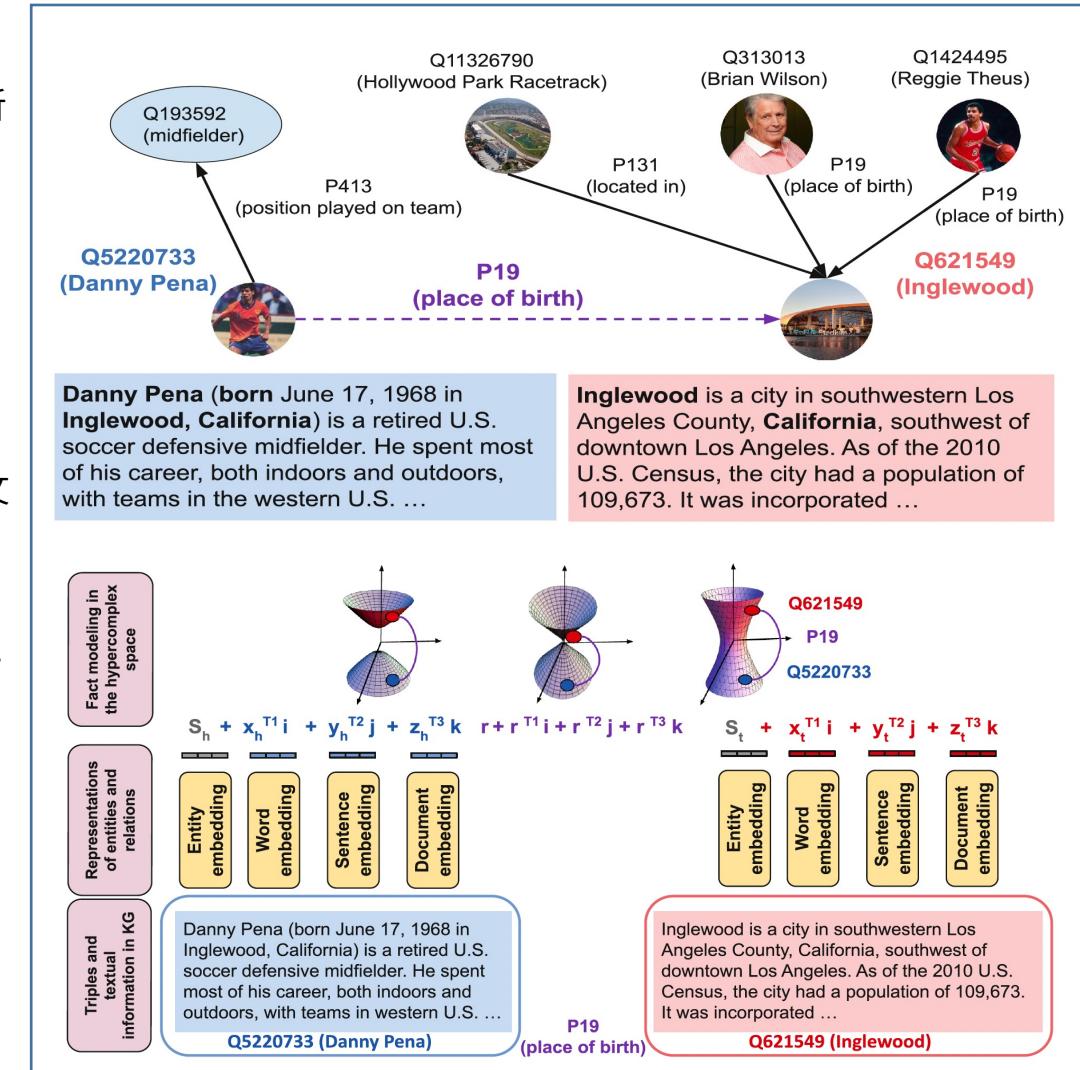
- 一言でいうと
コモンセンスとセンチメントの異種知識グラフによる感情分析(KSA)
- 動機
感情分析を行うLanguage Model (LM)を構築する際に1つの外部知識を導入する方法はあるが、複数の知識概念の獲得はまだできていない
- 手法
knowledge enhanced model for sentiment analysis(KSA)の提案した。
LM層、GNN層、および融合ユニットからなる複数の融合層により感情分類を行う。
GNN層はコモンセンスKGによりLM層は任意の事前学習済みモデルにより計算
- 背景
LMでは一般事象などには強いが、感情に関する知識が必要だった→KG
- 結果
4つのデータセットで検証。BERT, RoBERTaなどの言語モデルと感情データでファインチューニングしたモデルよりも精度向上を達成
- 考察
KG埋め込みや融合ユニットの有無精度を比較→ 複数のコンポーネントを組み合わせることが精度向上に寄与したと言える
- 課題
データ品質の問題や複雑性の増加



Integrating Knowledge Graph Embeddings and Pre-trained Language Models in Hypercomplex Spaces

Mojtaba Nayyeri, Zihao Wang, Mst. Mahfuja Akter, Mirza Mohtashim, Md Rashad Al Hasan Rony, Jens Lehmann and Steffen Staab/University of Stuttgart, Stuttgart, Germany

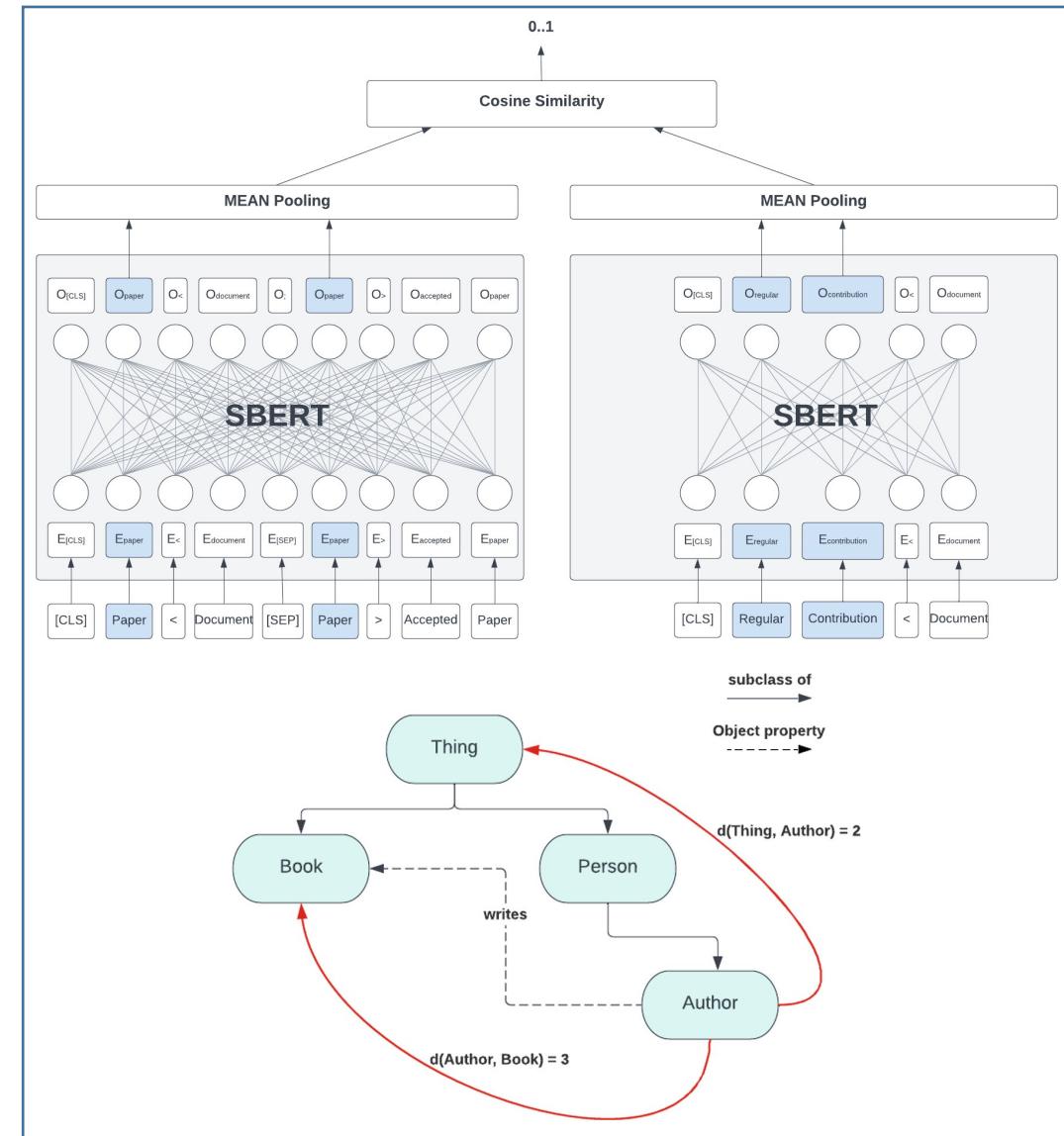
- 一言でいうと
知識グラフ埋め込みと複数の事前学習済み言語モデルを超複合空間で統合する新しいフレームワーク
- 動機
既存の知識グラフ埋め込み(KGE)モデルが单一の事前学習済み言語モデルに依存しており、異なるモデル間での相補的な活用がまだ
- 手法
超複合代数を使用して、構造的知識グラフの埋め込みと複数のテキスト表現間の相互作用をモデル化。Dihedron Modelを用いて、構造的知識、単語レベル、文レベル、ドキュメントレベルの4つの異なる表現を統合して複雑空間上に表現
- 背景
KGはAIシステムの中で重要な要素だが、実世界の事実と比べ不完全であるためテキストが必要だった
- 結果
リンク予測のタスクで多くの手法と比較し、SOTAを達成
- 考察
小規模なデータセットから大規模なデータセットで有効であることを示した。特にデータがスパースな場合に、テキスト情報の統合の有用性を確かめることができた
- 課題
複数ソースの情報を統合し、マルチホップKG補完シナリオや他のタスクにモデルを適用



SORBET: a Siamese Network for Ontology Embeddings Using a Distance-based Regression Loss and BERT

Francis Gosselin and Amal Zouaq/ LAMA-WeST Lab, Departement of Computer Engineering and Software Engineering, Polytechnique Montreal, 2500 Chem. de Polytechnique, Montréal, QC H3T 1J4, Canada

- 一言でいうと
BERTと距離ベースの回帰損失を使用してオントロジーの埋め込みを行うSiamese ネットワークであるSORBET
- 動機
オントロジー関連タスクのための表現学習方法が注目されているが、オントロジーの構造に忠実で意味的に関連性のあるオントロジーの埋め込みを構築するために大規模言語モデルを適応する研究は少ない
- 手法
概念構造を把握するTree Walkにより作成されたデータからSentence BERTが学習される。SBERTによりエンコードされた埋め込みはオントロジーのクラス間の意味的な距離を把握するために、距離ベースの損失関数と組み合わせて計算された
- 背景
KG埋め込みとオントロジー埋め込みは別々ですべきである
- 結果
複数のオントロジデータセットのサブキャプショントスクで最高性能を達成
- 考察
回帰損失を用いることでオントロジーでの距離を埋め込んでいる
- 課題
異なるルールの組み合わせを実験し、オントロジーの概念間の距離推定を改善



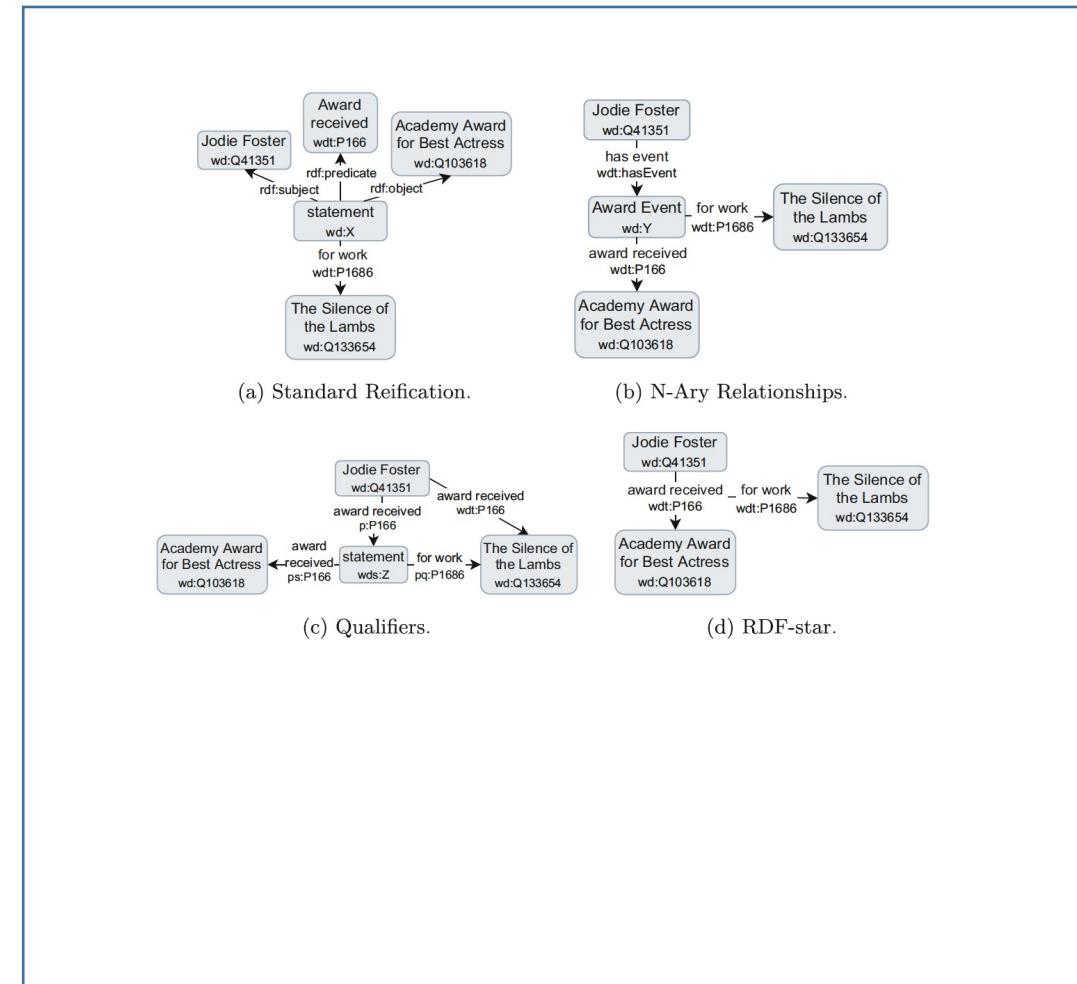
Session 9A: Knowledge Graph Embeddings IV

吉丸 直希

Comparison of Knowledge Graph Representations for Consumer Scenarios

Ana Iglesias-Molina, Kian Ahrabian, Filip Ilievski, Jay Pujara and Oscar Corcho/Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain

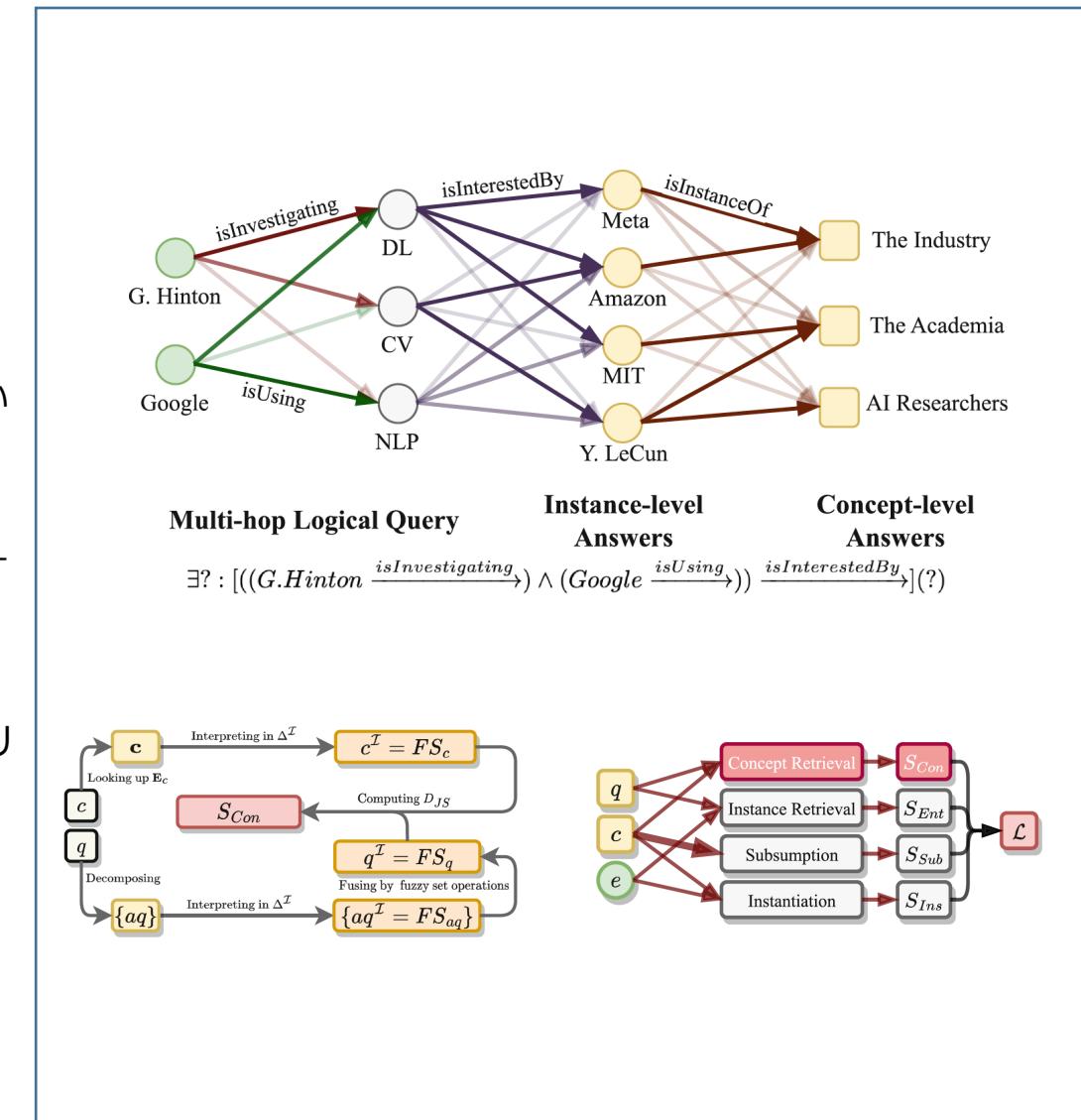
- 一言でいう
消費者シナリオにおける知識グラフ表現の適合性を分析
- 動機**
RDFが伝統的なモデル以外にも、プロパティグラフ、Wikidataモデル、RDF-starなど様々あるが、その表現が消費者シナリオ（知識探索、体系的クエリなど）にどう影響を与えるかを調査した研究
- 手法**
標準再化、N-ary関係、Wikidata修飾子、RDF-starの4つの表現方法を対象。知識探索タスクのためのユーザースタディ、合成データセットと実世界のデータセットを用いた体系的クエーリングの評価、知識グラフの埋め込みモデルを用いたグラフ完成タスクのパフォーマンス測定
- 背景**
知識グラフの使用の際に、どういうシナリオで何が最適化の一貫した見解がない
- 結果**
QualifiersとRDF-starが知識探索と体系的クエリに適する。標準再化モデルは高いパフォーマンスを出すがユーザにとって使いづらい可能性
- 考察**
標準再化（Standard Reification）は反直感的な構造で、ナビゲートする際に時間がかかるだけでなく、正確で完全な情報の取得が複雑
- 課題**
異なる知識グラフ表現の間の相互運用性を促進し、有用性を高める



Neural Multi-hop Logical Query Answering with Concept-level Answers

Zhenwei Tang, Shichao Pei, Xi Peng, Fuzhen Zhuang, Xiangliang Zhang and Robert Hoehndorf/ University of Toronto, Toronto, ON, Canada

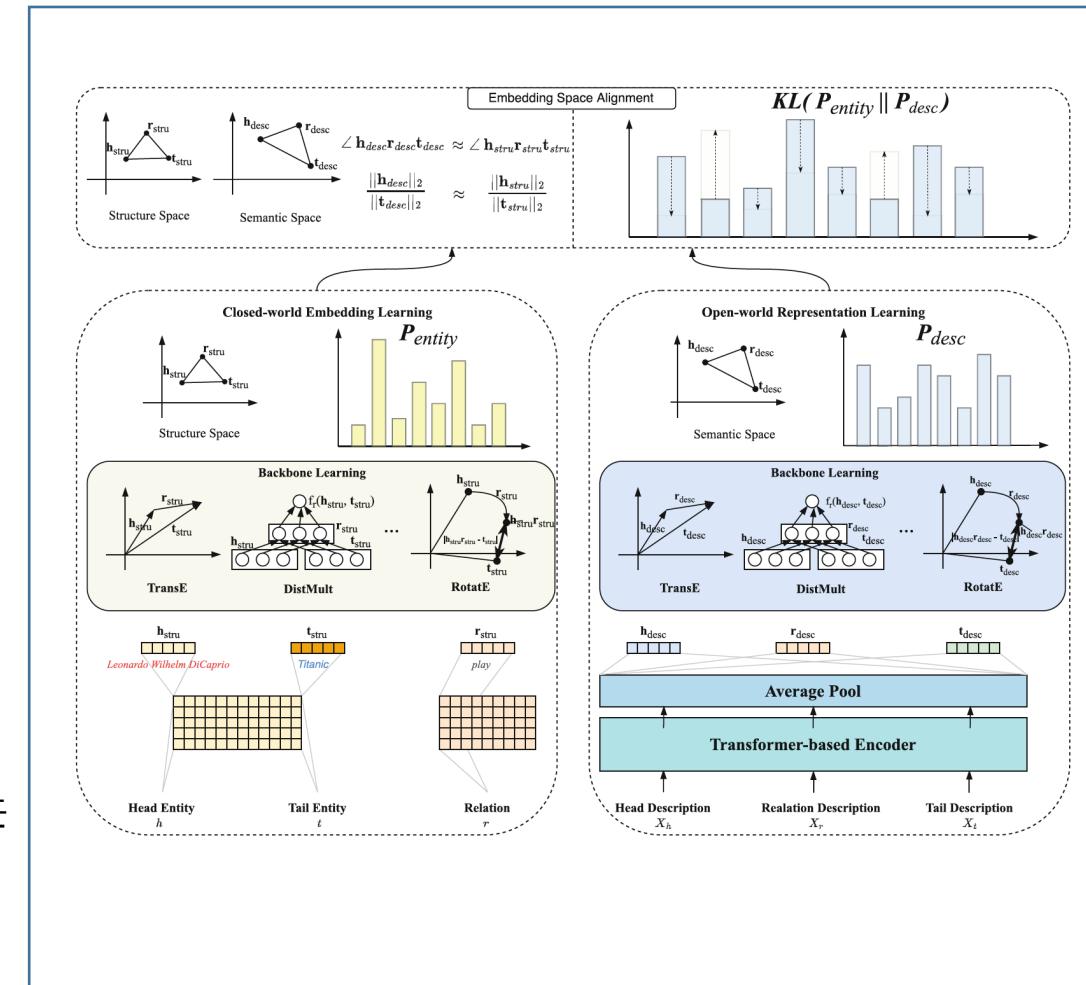
- 一言でいう
概念レベルの回答を提供するニューラルマルチホップ論理クエリ応答（LQAC）の問題を定式化
- 動機
論理クエリ応答（LQA）システムは、インスタンスレベルの回答のみを提供しており、ユーザーがより記述的な概念レベルの回答を求めるケースに対応できていなかった
- 手法
概念、インスタンス、クエリ間の関係性をモデル化するために、ファジ集合演算を用いた複数の演算子を設計
- 背景
関係データの探索において論理クエリ応答は基本だが、概念レベルの回答を提供する能力が不足
- 結果
複数の実世界のデータセットにおいて、概念レベルとインスタンスレベルの両方のクエリで従来のLQAシステムを上回る
- 考察
バイオメディカルなど特定の分野における知識発見におけるLQACの重要性を強調
- 課題
ALC記述論理における概念記述を含むより複雑なオントロジーに対する拡張、および否定クエリの取り扱い



ASKRL: An Aligned-Spatial Knowledge Representation Learning Framework for Open-world Knowledge Graph

Ziyu Shang, Peng Wang, Yuzhang Liu, Jiajun Liu and Wenjun Ke/ School of Computer Science and Engineering, Southeast University, Nanjing, China

- 一言でいうと
オープンワールド知識グラフにおけるゼロショットエンティティの問題に対処
- 動機
クローズドワールド知識表現学習(KRL)モデルは、ゼロショットエンティティを効果的に処理できず、オープンワールド設定での使用に限界
- 手法
SKRLは、構造化された埋め込み層、説明エンコーディング層、そして埋め込み空間整列層の3つの主要なコンポーネントで構成。エンティティとリレーションの表現を構造化された空間と意味空間の両方で学習
- 背景
オープンワールド知識グラフは、新たなエンティティやリレーションが継続的に追加されるため、既存のKRLモデルでは対応できない
- 結果
ASKRLはオープンワールド知識グラフ完了データセットにおいて、強力なベースラインモデルを一貫して上回る。特にBERTベースのエンコーダの際に顕著な向上
- 考察
KRLモデルやトランسفرベースのエンコーダーを使用することで、さらなる改善が期待
- 課題
異なる種類の知識グラフにおけるASKRLの適用性、多言語や他ドメインでの実験



6A: Knowledge Graph Embeddings II

小柳佑介（富士通株式会社）

Entity-Relation Distribution-aware Negative Sampling for Knowledge Graph Embedding

Naimeng Yao¹, Qing Liu², Yi Yang³, Weihua Li⁴, and Quan Bai¹/ ¹University of Tasmania, Hobart, Australia ²Data61, CSIRO, Hobart, Australia ³Hefei University of Technology, Hefei, China ⁴Auckland University of Technology, Auckland, New Zealand

- 一言でいうと

Entity-Relationペアに同じ数のネガティブサンプルを割り当てる影響を調査し、新しいnegative sampling手法を提案

- 動機

Entity-Relationペアの分布はロングテールであることが多く、各Entity-Relationペアに同じ数のネガティブサンプルを割り当てるることは問題。

- 手法

negative sampleの数の、学習への影響を調査。

調査に基づき、様々な数のnegative sampleを割り当てる、Entity-Relation分布を意識したネガティブサンプリング手法を提案。

- 背景

これまでのnegative samplingの研究は、不均衡な分布のERペアに同じ数のnegative samplesを割り当てる効果を検討しておらず、false negative samplesを最小化しながら質の高いnegative samplesを生成する課題に取り組んでいない。

- 結果

従来のKGEおよびNN-based KGEモデルで試験して検証。
提案手法はSOTAより優。

- 考察

実験結果から、NN-based モデルにおいては、グローバルな特徴の組み込みが訓練プロセスにおけるnegative sampleの有効性に影響。

- 課題

NNベースモデルへのアプローチの有効性を高める方法をさらに探求。

Table 3. Performance comparison on DistMult, ComplEx, TransE and RotatE.

Dataset	KGE model	DistMult				ComplEx			
		Method	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@10
FB15K237	KBGAN	26.7	-	-	45.3	28.2	-	-	45.4
	NSCaching	28.3	-	-	45.6	30.2	-	-	48.1
	unif	21.9	13.1	23.4	41.1	22.7	13.7	24.5	42.1
	Bernoulli	25.9	17.4	28.3	43.3	31.0	21.2	34.3	51.3
	Self-Adv	30.9	22.1	33.6	48.4	32.2	23.0	35.1	51.0
	Subsampling	29.9	21.2	32.7	47.5	32.8	23.6	36.1	51.2
WN18RR	ERDNS	35.0	25.7	38.3	53.7	35.7	26.2	39.3	54.7
	KBGAN	38.5	-	-	44.3	42.9	-	-	47.0
	NSCaching	41.3	-	-	45.5	44.6	-	-	50.9
	unif	42.7	39.0	43.7	50.7	44.8	41.5	46.4	51.7
	Bernoulli	40.0	37.3	40.7	45.2	46.0	42.4	47.8	53.0
	Self-Adv	43.9	39.4	45.2	53.8	47.1	42.8	48.9	55.7
YAGO3-10	Subsampling	44.6	40.0	45.9	54.4	47.6	43.3	49.3	56.3
	ERDNS	45.5	41.1	46.7	54.6	48.4	44.3	50.0	56.4
	NSCaching	40.3	-	-	56.6	40.5	-	-	57.8
	unif	45.4	35.7	51.0	64.5	46.4	36.6	51.9	64.8
	Bernoulli	49.0	39.4	54.2	67.2	50.3	40.9	55.5	68.0
	Self-Adv	44.8	34.7	50.4	64.3	47.3	37.4	53.2	66.1
Dataset	Subsampling	45.9	34.6	52.7	66.2	49.1	38.1	56.0	68.5
	ERDNS	52.6	43.7	57.8	69.4	54.6	45.9	59.8	70.3
	KGE model	TransE				RotatE			
	Method	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
FB15K237	KBGAN	29.4	-	-	46.7	-	-	-	-
	NSCaching	30.0	-	-	47.4	30.9	22.1	34.1	48.6
	unif	31.4	21.5	35.2	51.5	31.9	22.3	35.3	51.7
	Bernoulli	33.4	23.9	37.0	52.6	33.6	24.1	37.1	52.7
	Self-Adv	32.9	23.0	36.8	52.7	33.6	23.9	37.4	53.0
	Subsampling	33.6	24.0	37.3	52.9	34.0	24.5	37.6	53.2
WN18RR	ERDNS	34.2	24.8	37.9	53.0	34.5	25.2	37.9	53.7
	KBGAN	18.6	-	-	45.4	-	-	-	-
	NSCaching	20.0	-	-	47.8	47.1	42.9	48.9	55.3
	unif	22.4	1.5	40.1	51.9	47.1	42.8	48.6	55.5
	Bernoulli	23.2	2.2	41.4	52.2	47.5	43.1	49.1	56.0
	Self-Adv	22.3	1.3	40.1	52.8	47.6	43.1	49.5	57.3
YAGO3-10	Subsampling	23.0	1.9	40.7	53.7	47.8	42.9	49.8	57.4
	ERDNS	23.8	3.0	41.8	52.8	47.9	43.5	49.4	56.5
	NSCaching	30.7	-	-	50.7	42.3	33.8	47.1	59.7
	unif	48.2	37.4	55.1	67.3	49.8	40.2	55.2	67.7
	Bernoulli	50.3	40.6	56.3	67.9	50.2	40.8	55.8	67.9
	Self-Adv	51.2	41.5	57.6	68.3	50.8	41.8	56.5	67.6
Dataset	Subsampling	51.3	41.9	57.2	68.1	51.0	41.9	56.5	67.8
	ERDNS	51.7	42.2	57.6	68.7	51.8	42.7	57.0	68.8

Table 4. Performance comparison on ConvE and CompGCN.

KGE model	Dataset	FB15K237				WN18RR				
		Method	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@10	
ConvE	ConvE	unif	31.5	22.6	34.4	49.8	41.0	36.9	42.4	49.0
		Bernoulli	29.8	21.4	32.5	46.9	39.7	35.9	41.0	47.9
		Self-Adv	31.9	22.7	34.8	50.4	40.3	37.0	41.4	47.0
		ERDNS	32.3	23.4	35.4	50.2	41.3	38.0	42.6	48.4
CompGCN (TransE)	CompGCN (TransE)	unif	32.9	24.1	36.1	50.1	18.9	5.6	25.6	47.5
		Self-Adv	28.5	20.0	31.2	45.7	20.4	5.3	29.6	49.3
		ERDNS	33.6	26.2	36.2	50.9	24.5	4.9	39.0	53.1
		unif	33.3	24.4	36.4	51.2	42.0	37.3	43.3	52.1
CompGCN (DistMult)	CompGCN (DistMult)	unif	29.2	21.0	31.8	45.9	43.0	39.2	43.1	51.8
		Self-Adv	34.2	25.2	37.4	52.3	43.4	39.2	44.3	52.5
		ERDNS	35.2	26.2	38.5	53.1	46.6	43.4	47.7	53.0
		unif	30.3	22.6	32.9	46.2	43.7	40.7	44.7	49.4
CompGCN (ConvE)	CompGCN (ConvE)	unif	35.2	26.2	38.5	53.1	46.6	43.4	47.7	53.0
		Self-Adv	30.3	22.6	32.9	46.2	43.7	40.7	44.7	49.4
		ERDNS	35.3	26.3	38.9	53.6	46.8	43.8	47.8	53.1

Negative Sampling with Adaptive Denoising Mixup for Knowledge Graph Embedding

Xiangnan Chen, Wen Zhang, Zhen Yao, Mingyang Chen, and Siliang Tang/Zhejiang University, Hangzhou, China

- 一言でいうと

DeMix: KGEのための、ネガティブサンプリングトリプルのノイズ除去手法の提案。

- 動機

ほとんどの既存手法は、存在しないトリプルがnegative トリプルであると仮定する（閉世界仮説）が、この方法ではノイズが含まれる可能性がある。例えば、KGが不完全で、真の事実が存在しない可能性がある。

- 手法

negative sampling トリプルを自己監視方式で判定しつつ高品質なnegativeトリプルを生成する、簡単に接続可能なノイズ除去mixup手法 DeMix を提案

- Marginal Pseudo-Negative Triple Estimator (MPNE): KGEモデル自身の結果を基に、擬似negativeトリプルと真negativeトリプルに分割。
- Adaptive Mixup (AdaMix): 適切な mix-up パートナーを選択し、Embedding空間上で混合して、partially positive トリプルやharder negative トリプルを生成。

- 背景

過去のnegative samplingは検索ベースの方法であり、非効率的。CANSは外部情報収集に手作業が必要であり、Bernoulliはサンプリングスキームが固定。

既存手法 MixKG は、KGEのためのhard negative トリプルの生成のために、既存のMixUp手法を活用しているが、提案手法ではノイズの多いnegativeトリプルを動的に洗練。

- 結果・考察

KG Completionタスクにおいて、DeMixが他手法よりも優。（WN18RRでのRotateEだと、指標によってはRW-SANSが優だが同程度か）

他のnegative sampling (Uniform, RW-SANS) と組合せることで、有意な改善。

Ablation Studyにより、各モジュールの有用性を検証。

- 課題

将来的に、トレーニングセット内の見えないパターンを持つ雑音のあるトリプルを認識するように拡張し、この方法にアクティブラーニングを適用

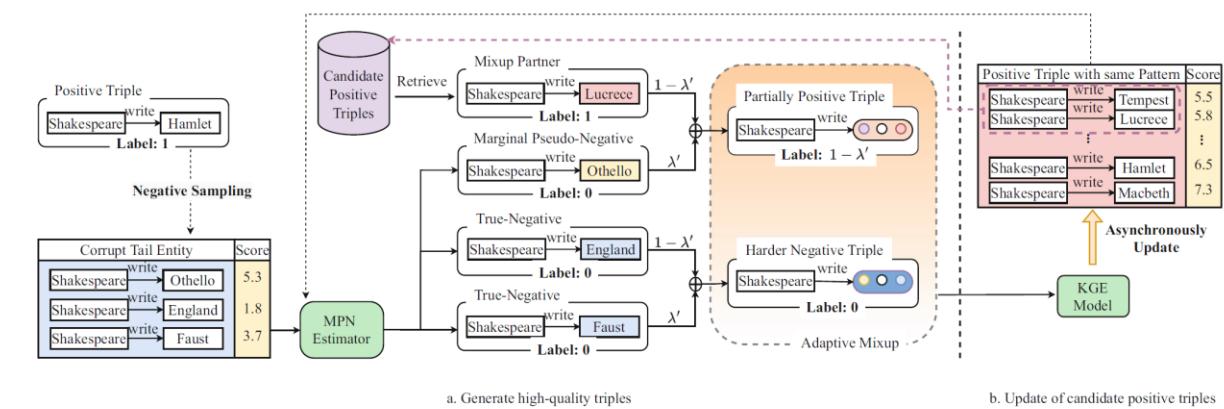


Fig. 2. An overview of the DeMix framework based on corrupting tail entity.

Biomedical Knowledge Graph Embeddings with Negative Statements

Rita T. Sousa¹, Sara Silva¹, Heiko Paulheim², and Catia Pesquita¹/

¹LASIGE, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal ²Data and Web Science Group, Universität Mannheim, Mannheim, Germany

- 一言でいうと

TrueWalks: KG表現学習プロセスにnegative statementsを組み込むアプローチを提案。

- 動機

negative statementsの考慮は、エンティティ要約や質問応答などのタスクや、タンパク質機能予測などのドメイン固有のタスクのパフォーマンスを向上させるが、これまでのKG Embeddingアプローチでは、negative statementsの探索に注意が払われていない。

- 手法

知識グラフ表現学習プロセスにnegative statementsを組み込む新しいアプローチ、TrueWalksを提案。

statementの種類ごとに2つのembeddingを一つずつ生成し、最後の潜在表現を得る。エンティティに欠けている機能に焦点を当てる。通常は、従来のskip-gramでembeddngsを学習(TrueWalks)。

グラフウォーク内のエンティティの順序に敏感な、structured skip-gramで学習するバリエーション(TrueWalksQA)もある。

- 背景

KGのnegative statementsにより、様々なアプリケーションにおいてパフォーマンス向上。

近年のKG Embeddingアプローチは、KGとオントロジーの意味的、構造的、字句的側面を考慮して表現を調整。OWL2Vec*は、逆経路横断を可能とする逆公理の宣言を考慮。このオプションはGene Ontologyには欠けている。

- 結果・考察

関係予測タスク（PPI予測、GDA予測）によりTrueWalksを評価。代表的なSOTAよりも優。

- 課題

疾患に関連する表現型の予測や鑑別診断など、negative statementsが決定的な役割を果たす他の生物医学アプリケーションへの一般化。

opposite statementsがエンティティの非類似性にどのように影響を与えるかを検討するため、language embeddingで提案されているようなcounter-fittingアプローチを探求。

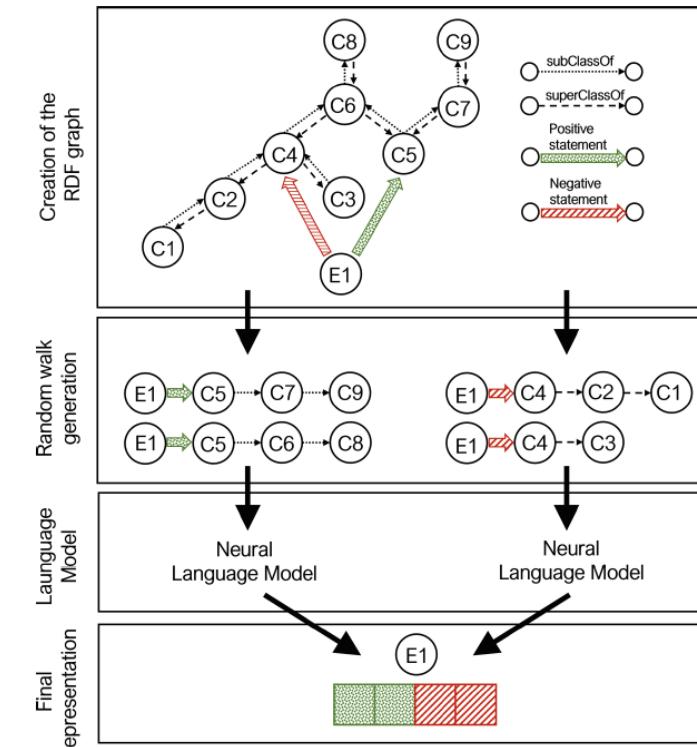


Fig. 3. Overview of the TrueWalks method with the four main steps: (i) creation of the RDF graph, (ii) random walk generation with negative statements; (iii) neural language models, and (iv) final representation.

8A: Knowledge Graph Embeddings III

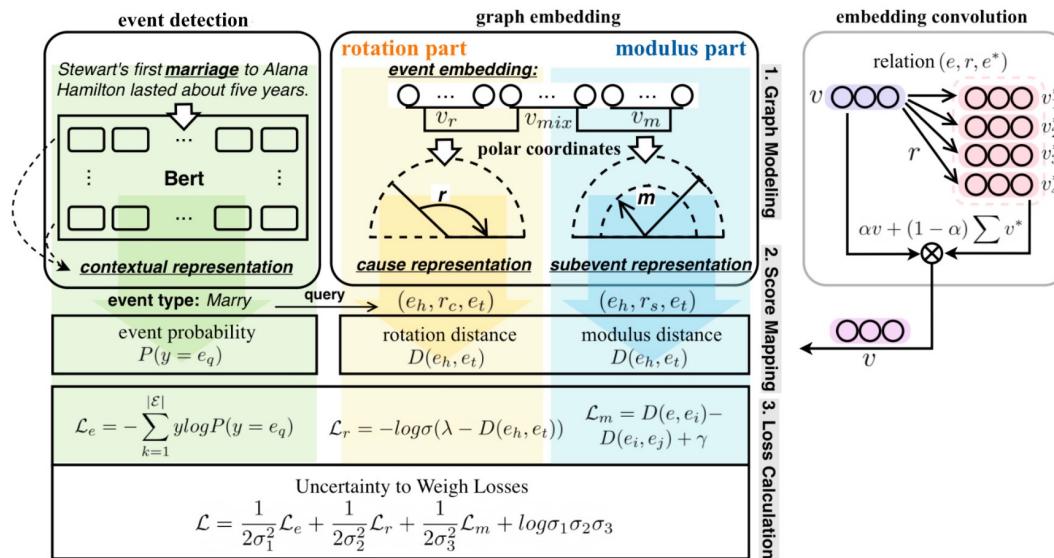
7C: Temporal Reasoning

産業技術総合研究所 江上周作

HAEEvent: Low-Resource Event Detection with Hierarchy-Aware Event Graph Embeddings

Guoxuan Ding^{1,2}, Xiaobo Guo¹, Gaode Chen^{1,2}, Lei Wang¹ and Daren Zha¹ / ¹Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China. ²School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

- 一言でいうと
テキストから構造化されたイベント情報を抽出するため、イベントグラフ埋め込みによる階層認識モデルHAEEventを提案
- 動機
深い意味的な関連性を捉えてイベント検出モデルの性能を向上させる
- 手法
原因-サブイベントの関係とイベントグラフ埋め込みを組み合わせた新たな階層認識モデルを提案。イベントを極座標に写像し、イベントペアを円上の回転で表現する回転ベースのアプローチを採用。
- 背景
イベント検出はデータ不足により限られたデータで学習しなければならないことが課題
抽象的なイベント関係を追加知識として利用することが有望視されている
- 結果
リソースの少ないイベント検出タスクにおいてより高い性能を達成
- 考察
a)原因表現、b)サブイベント表現、c)Ablation studyを含む様々な観点から分析して知見を説明
- 課題
より多くのイベント関係やより複雑な構造、他の情報抽出タスクへの拡張



A Comprehensive Study on Knowledge Graph Embedding over Relational Patterns Based on Rule Learning

Long Jin¹, Zhen Yao¹, Mingyang Chen², Huajun Chen^{2,3}, and Wen Zhang¹ / ¹School of Software Technology, Zhejiang University, Hangzhou, China. ²College of Computer Science and Technology, Zhejiang University, Hangzhou, ³China. Donghai laboratory, Hangzhou, China

- 一言でいう
ナレッジグラフ埋め込み(KGE)における関係パターンの定量的な分析と、追加学習無しでKGEの性能を向上させる手法の提案

• 動機

関係パターンに対するKGEモデルの包括的な定量的分析が行われていない

• 手法

4つの一般的な関係パターンに対する7つのKGEモデルの性能を2つのベンチマークで評価
関係パターンとKGEスコアの情報を組み合わせてモデルのスコア関数を変更(SPA)

• 背景

KGEの評価において関係パターンは重要な指標であり徐々に理解が深まっている
N対N関係、階層、等価、反転、対称、合成、…

• 結果

理論上特定の関係パターンをサポートしているKGEでも、そうでないKGEに対する優位性を保証しない。エンティティ頻度が関係パターンの性能に与える影響は異なる。有意に優れたKGEモデルは、すべての関係パターンに一貫して優れている。

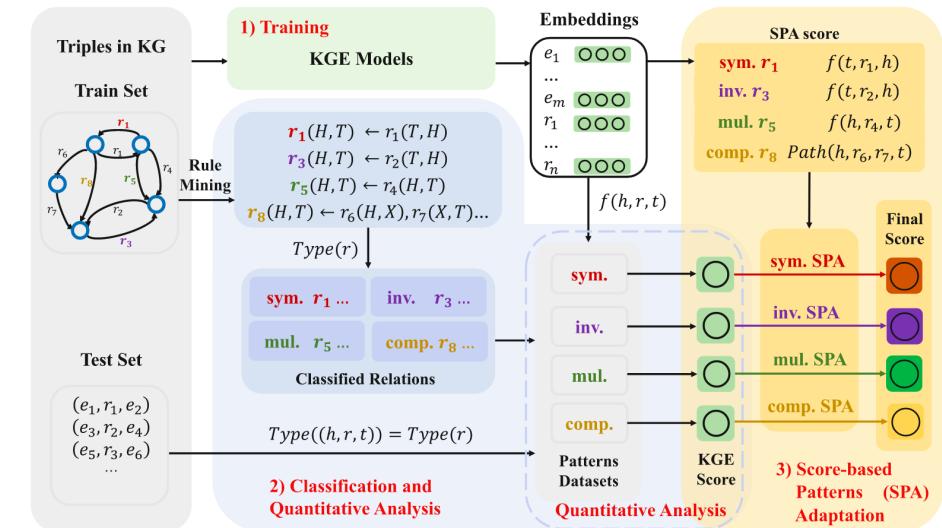
FB15K237とWN18RRの様々なパターンデータセットにおいて、SPAによる精度向上が見られた

• 考察

直感と事実が異なる原因の一つは関係間の関連性の複雑さ。関係パターンが識別にくくなっていると考えられる

• 課題

マクロな視点からの関係全体の相関分析、ネガティブサンプリング、損失関数にもっと注意を払うべき



FeaBI: A Feature Selection-Based Framework for Interpreting KG Embeddings

Youmna Ismaeil^{1,2}, Daria Stepanova¹, Trung-Kien Tran¹, and Hendrik Blockeel² / ¹Bosch Center for Artificial Intelligence, Renningen, Germany. ²KU Leuven, Leuven, Belgium

- 一言でいうと
KGEに対して解釈可能なベクトルを生成する手法

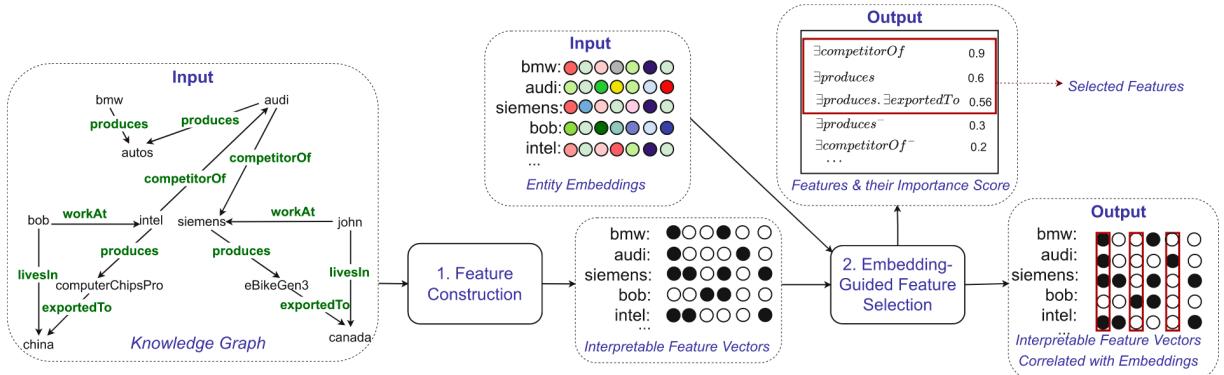
- 動機
KGEはブラックボックス

- 手法
KGから命題特徴を抽出し記述論理で表現し、エンティティの近傍に依存するブーリアンベクトル（特徴ベクトル）を構築。KGEモデルにおけるエンティティ表現を再構成するため、回帰ランダムフォレストを使用して特徴をランク付け。

- 背景
パス生成などの既存研究はKGEの結果を説明することに重点を置いているが、提案手法はKGEの振る舞いを模倣するように解釈可能な特徴表現を生成

- 結果
ベースライン（乱数ベクトル）と比較してKGE（INK, TransE, CompGCN, NodePiece, Snore）ではランダムフォレストのMSEが低く、入力特徴ベクトルと対応するKG埋め込みとの間の意味のある関係を識別できることを立証

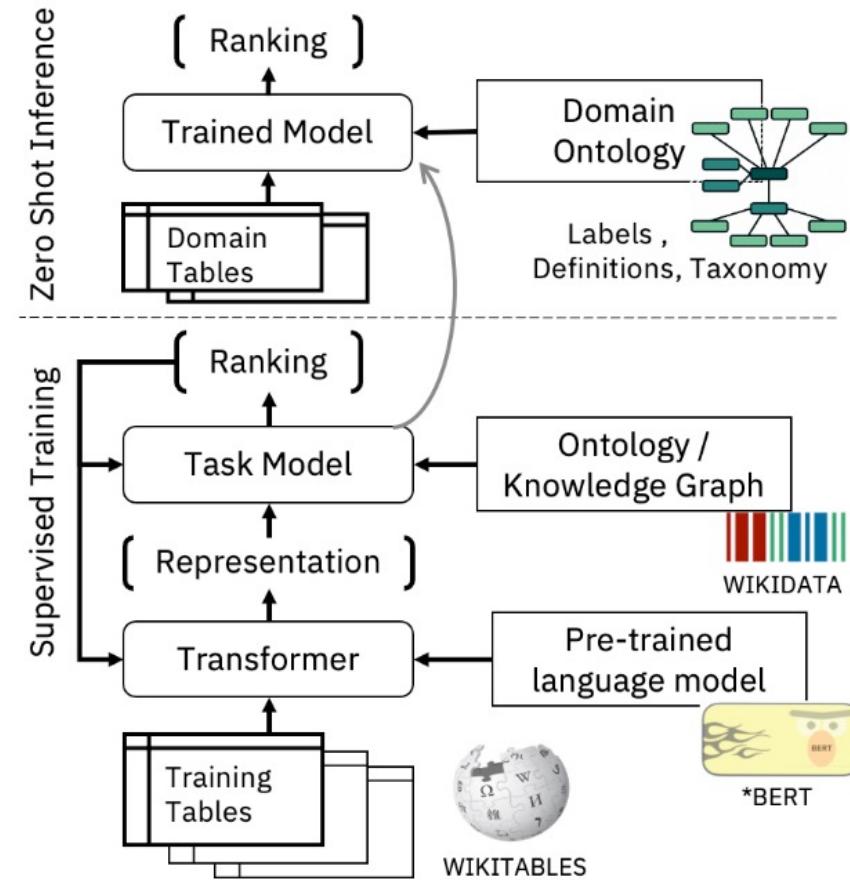
- 考察
考察はあるが読めていない
- 課題
KGが大きくなると、特徴選択アルゴリズムの探索空間が増大し、スケーラビリティの問題につながる



Linking Tabular Columns to Unseen Ontologies

Sarthak Dash, Sugato Bagchi, Nandana Mihindukulasooriya and Alfio Massimiliano Gliozzo / IBM Research AI, Yorktown Heights, NY, USA

- 一言でいうと
テーブルの列を未知のオントロジーのタイプにリンクするための手法
- 動機
データレイクにおけるメタデータにはデータベースの名前、テーブル、カラム、関連するデータベーススキーマが含まれており、外部オントロジーへのマッピングによる情報の発見、拡張、可視化をしたい
- 手法
既存のオントロジーへのマッピングが既存のテーブルに対してTransformerベースの深層学習モデルを学習させ、そのモデルを未知オントロジーに適用する
- 背景
現在のアプローチはルールベースか同じオントロジー内での学習が必要な手法であり、データ消費者の役割ごとに異なるオントロジーに列をリンクするケースでは現実的ではない
- 結果
3つの新しいデータセットを導入。様々な評価と分析を通して提案アプローチの有効性を実証。
- 考察
考察はあるが読めていない
- 課題
訓練時に使用したドメインとかけ離れたドメインに適用する場合。2D関係テーブル以外の場合。



FORECASTTKGQUESTIONS: A Benchmark for Temporal Question Answering and Forecasting over Temporal Knowledge Graphs

Zifeng Ding^{1,2}, Zongyue Li^{1,3}, Ruoxia Qi¹, Jingpei Wu¹, Bailan He^{1,2}, Yunpu Ma^{1,2}, Zhao Meng⁴, Shuo Chen^{1,2}, Ruotong Liao^{1,2}, Zhen Han¹, and Volker Tresp¹ / ¹LMU Munich, Germany. ²Siemens AG, Germany. ³Munich Center for Machine Learning (MCML), Germany. ⁴ETH Zurich, Switzerland

- 一言でいう
未来についての質問に答える時間的KG質問応答(TKGQA)タスクのベンチマークデータセットForecastTKGQuestionsと予測TKGQAタスクの提案

- 動機
最近のTKG補完は未来のタイムスタンプでの予測に注目

- 手法
エンティティ予測質問、Yes-unknown質問、ファクト推論質問の三種類を含む
TKG予測モデルTANGOとBERTを採用した提案モデルForcastTKGQA

- 背景
既存のTKGQA研究は未来の事実の予測を対象としていない

- 結果
提案モデルがSOTA。TKGQA予測においてはTKG予測モデルがTKG補完モデルよりも有効？→有効。データセットは解答可能？→GTのTKG情報があれば可。データセットは効率的？→データ増加するほど精度が向上するため効率的

- 考察
人間と比較するとまだ改善の余地が大きい。

- 課題
質問タイムスタンプ t_q におけるGTのTKG情報 G_{t_q} を正確に推論。効果的なマルチホップ推論。より良いファクト推論のためのTKGQAモデル開発。

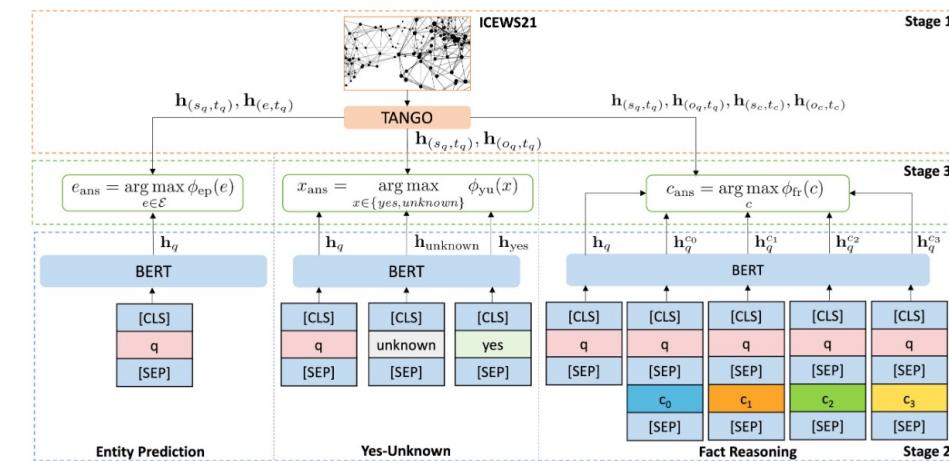


Fig. 2. Model structure of FORECASTTKGQA.

TEMPORALFC: A Temporal Fact Checking Approach over Knowledge Graphs

Umair Qudus¹, Michael Röder¹, Sabrina Kirrane², and Axel-Cyrille Ngonga Ngomo¹ / ¹DICE Group, Department of Computer Science, Universität Paderborn, Germany. ²Institute for Information Systems and New Media, Vienna University of Economics and Business, Austria

- 一言でいうと
与えられたKGのアサーションの正しさと時間的妥当性を評価する時間的ファクトチェック
- 動機
ほとんどのファクトチェックアプローチはアサーションが特定の時間間隔でのみ有効であるという事実を考慮しない
- 手法
TKGの事前学習埋め込みを転移学習を利用するニューラルネットワークベースのアプローチを提案。トリプルの検証だけでなく真であった年も予測。
- 背景
過去10年間KGのファクトチェックングの研究がされている。非構造化情報を利用するもの、構造化情報を利用するもの、ハイブリッドの3種に大別できる
- 結果
提案モデルがSOTA
- 考察
他のほとんどのアプローチが時間情報を考慮していないため差がついた。他のアプローチと比較して時間を要するのは時間的埋め込みの入力ベクトルが大きいため
- 課題
Time periodのサポート

