

人工知能学会第56回SWO研究会 × ナレッジグラフ若手の会
ISWC2021サーベイ

2022年3月11日 オンライン

企画の趣旨

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

ISWC2021情報源

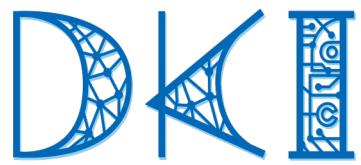
- プログラム
 - <https://iswc2021.semanticweb.org/final-programme>
- 論文
 - ISWC 2021, LNCS 12922: <https://link.springer.com/book/10.1007%2F978-3-030-88361-4>
 - Posters, Demos and Industry Track: <http://ceur-ws.org/Vol-2980/>
- 受賞情報
 - https://twitter.com/iswc_conf

今回の新たな取組

- サーベイ用テンプレートを用意し、こちらの項目を意識したサーベイを推奨（任意）
 - <https://github.com/knowledgegraph-yra/Survey>
- 発表終了後、有志にて皆様のサーベイ内容紹介スライドを上記テンプレートに基づいて構造化しアーカイブする予定
 - 発表者ご自身で入力いただけますと大変助かります。
 - すべての項目が埋まらず不完全なままでも問題ございません。
- ご協力いただける方は、下記よりIssuesへの投稿をお願いいたします。
 - <https://github.com/knowledgegraph-yra/Survey/issues>

担当者一覧

発表順	担当者	担当セッション
1	江上周作	Session 2A: Visual Models Session 5C: Semantic Matching
2	藤原浩司	Session 3B: Data Transformation Session 4C: Data Analytics
3	長野伸一	Session 3D: Industry 3 Session 2D: Industry 2
4	砂川英一	Session 4A: Ontologies Session 1D: Industry 1
5	山本泰智	Session 5A: Linked Data Applications Session 5B: Validation
6	鵜飼孝典	Session 2B: Learning from Wikidata Session 1A: Question Answering
7	キャンセル	Session 1C: Logics and Reasoning Session 1B: Society and Semantic Technologies
8	小柳祐介	Session 4B: Information Extraction Session 2C: Semantic Integration
9	光石豊	Session 3A: Knowledge Graph Embeddings



ISWC2021 サーベイ

Session 2A: Visual Models

Session 5C: Semantic Matching

産業技術総合研究所
江上周作

Graphhopper: Multi-hop Scene Graph Reasoning for Visual Question Answering

- Rajat Koner¹, Hang Li^{1,2}, Marcel Hildebrandt^{1,2}, Deepan Das³, Volker Tresp^{1,2}, and Stephan Günnemann³
 - ¹Ludwig Maximilian University of Munich, Munich, Germany. ²Siemens AG, Munich, Germany. ³Technical University of Munich, Munich, Germany

- 一言でいうと
 - シーングラフと質問文の埋め込みと強化学習により、シーングラフ上のパス探索をするVisual Question Answering (VQA)手法「Graphhopper」を提案
- 動機
 - VQAは、画像と自然言語の2つのモダリティを理解し推論する、野心的なタスクである。
 - VQAの技術と、ナレッジグラフにおける統計的関係学習の最近の手法を組み合わせることを目的とする。
- 手法
 - 正解に対応するノードまでのパス探索に強化学習を採用
 - 質問文をTransformerでベクトル空間に埋め込み、シーングラフをGloVeとGraph Attention Network(GAT)で埋め込む
 - 各埋め込みを基にLSTMによりエージェントの行動履歴を埋め込む
 - 方策ネットワークのLSTMにより次の行動を決定する
- 結果
 - 既存手法を大幅に上回るパフォーマンスを達成

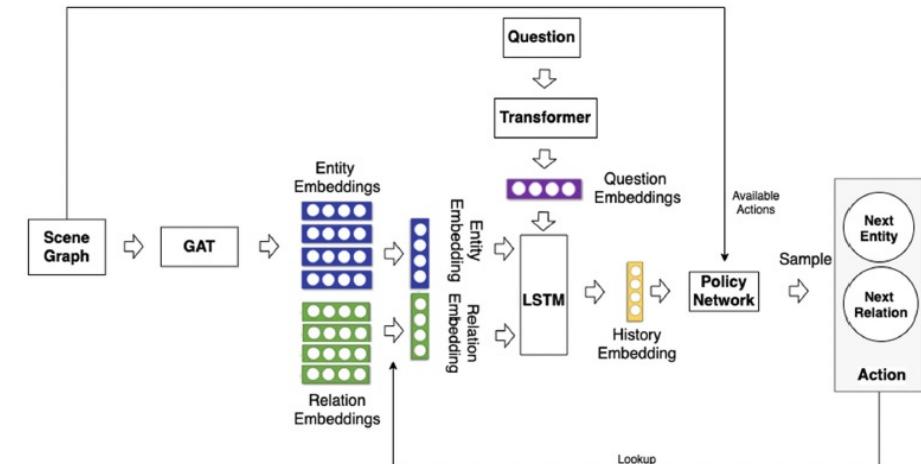


Fig. 2. The architecture of our scene graph reasoning module.

Learning Visual Models Using a Knowledge Graph as a Trainer

- Sebastian Monka^{1,2}, Kavdim Halilaj¹, Stefan Schmid¹, and Achim Rettinger²

- ¹Bosch Research, Renningen, Germany. ²Trier University, Trier, Germany

- 一言でいうと

- ドメイン不变な補助知識を使用して学習をSuperviseするNeuro-symbolicアプローチ「KG-NN」を提案する。

- 動機

- コンピュータビジョンのアプローチは学習する画像データの分布に依存しているため、ソースドメインと異なるターゲットドメインに応用すると失敗する傾向にある。
 - よりロバストに学習してドメインシフトするための手法を提案する。

- 手法

- ナレッジグラフ (KG) によって獲得されるドメイン不变の事前知識を使うNeuro-symbolicアプローチ
 - ニューラルネットワーク (NN) により学習されるビジュアル埋め込みと知識グラフ埋め込みを組み合わせるために対照損失関数を適用

- 結果

- WordNetのサブセットとMini-ImageNetを用いたドメイン汎化実験の結果、GloVeなどの既存手法より高い精度を実現。
 - 道路標識認識のための少量のKG(RSKG)を用意して埋め込み、The German Traffic Sign Dataset (GTSRB)とChinese Traffic Sign Dataset (CTSD)で実験した結果、既存手法と比較して高い精度を実現。

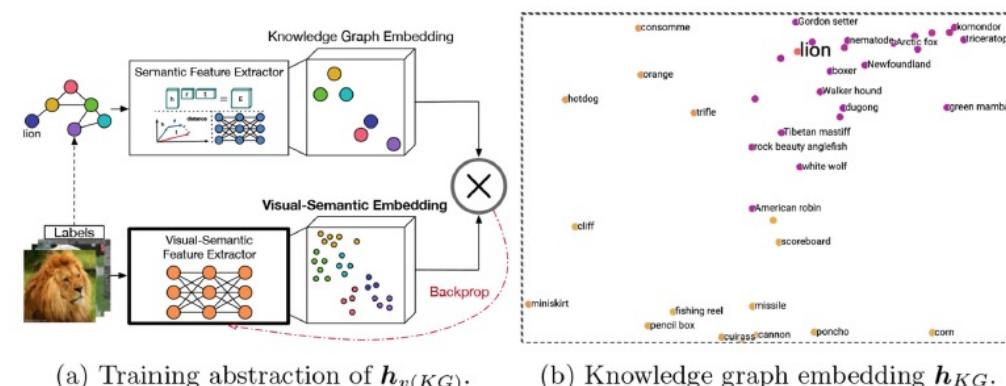


Fig. 2. KG-NN Approach: a) the main building blocks for learning a visual-semantic embedding space $h_{v(KG)}$ using a knowledge graph as a trainer; b) the 2D projection of the semantic-embedding h_{KG} represented in a knowledge graph.

Zero-Shot Visual Question Answering Using Knowledge Graph

- Zhuo Chen^{1,2}, Jiaoyan Chen³, Yuxia Geng^{1,2}, Jeff Z. Pan⁴, Zonggang Yuan⁵, and Huajun Chen^{1,2}

¹College of Computer Science and Hangzhou Innovation Center, Zhejiang University, Hangzhou, China, ²AZFT Joint Lab for Knowledge Engine, Hangzhou, China, ³Department of Computer Science, University of Oxford, Oxford, UK, ⁴School of Informatics, The University of Edinburgh, Edinburgh, UK, ⁵NAIE CTO Office, Huawei Technologies Co., Ltd., Shenzhen, China

- 一言でいうと
 - ナレッジグラフとマスクベースの学習機構を用いたZero-ShotのVQAアルゴリズムの提案
- 動機
 - 現在、VQAは答えが直接的に画像に含まれていないシーンの理解に関しては外部知識に依存している。外部知識を組み込む既存手法はパイプラインのアプローチを採用しているが、このようなアプローチは一部がうまく機能しないときに全体のパフォーマンスの低下につながる。
 - また、既存のアプローチの大半は、実世界応用における答えのバイアス問題を考慮していない。
- 手法
 - ナレッジグラフ (KG) とマスクベースの学習機構を用いたZS-VQAアルゴリズムの提案
 - Unseenな答えに対してZS-VQAを評価する Zero-shot Fact VQA datasetの提案
- 結果
 - 実験の結果SOTAを達成し、さらに通常のF-VQAタスクで既存のend-to-endモデルを劇的に強化できることを裏付けた。

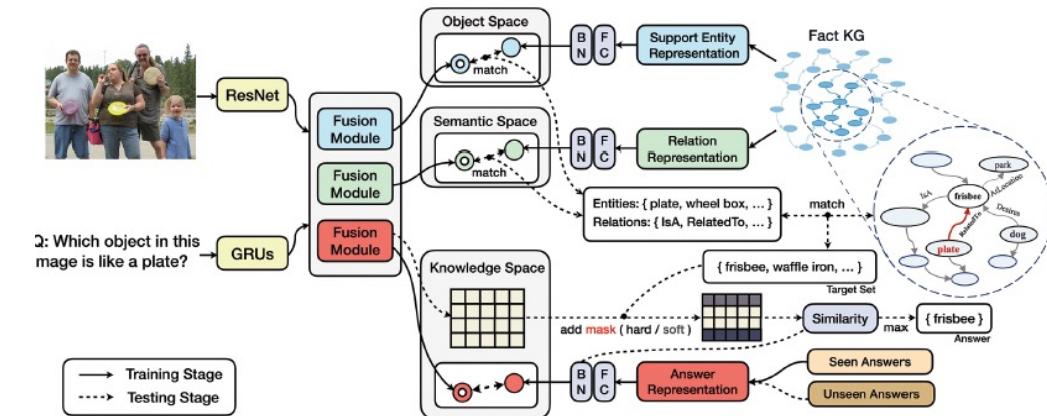


Fig. 2. An overview of our framework.

Graph-Boosted Active Learning for Multi-source Entity Resolution

- Anna Primpeli¹ and Christian Bizer¹
 - ¹Data and Web Science Group, University of Mannheim, Mannheim, Germany
- 一言でいうと
 - マルチソースのエンティティ解決のためのActive Learning手法であるALMSERを提案。
- 動機
 - エンティティ解決のための既存のactive learning手法は2つのソースのマッチングを対照としており、Web of Dataなどのマルチソースの設定にのみ存在する信号を無視している。
- 手法
 - マルチソースエンティティ解決のためのgraph-boosted active learning手法であるALMSERを提案。
 - グラフの推移性や最小カットなどのグラフ信号を活用して、学習器の予測から潜在的にfalse negativeとfalse positiveのレコードペアを見つける。
- 結果
 - グラフ信号を活用することで、全てのタスクのF1スコアについて、マージンベースおよびquery-by-committee戦略を使用したactive learningよりも良い結果を得た

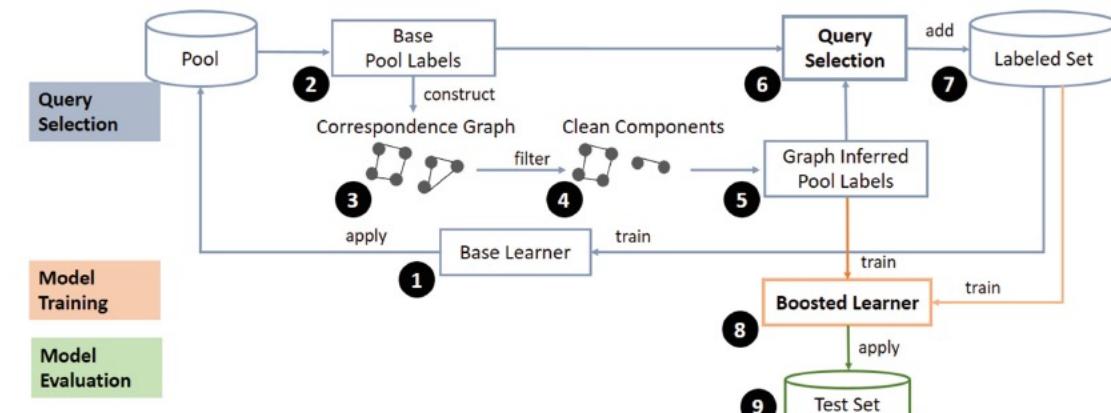


Fig. 2. Overview of the ALMSER algorithm.

Towards Neural Schema Alignment for OpenStreetMap and Knowledge Graphs

- Alishiba Dsouza¹, Nicolas Tempelmeier², and Elena Demidova¹

• ¹Data Science and Intelligent Systems (DSIS) University of Bonn, Bonn, Germany. ²L3S Research Center, Leibniz Universität Hannover, Hannover, Germany

• 一言でいうと

- OpenStreetMapのタグをWikidata等のナレッジグラフ上のクラスにリンクするNeural Class Alignment (NCA)モデルの提案

• 動機

- 豊富な地理的エンティティを含むOpenStreet Map(OSM)とナレッジグラフ(KG)とのアライメントをしたい。

• 手法

- OSMのタグとKG上で対応するクラスをリンクする、インスタンスベースのニューラルアプローチNeural Class Alignment (NCA)モデルを提案

- KGとOSMからの特徴空間をジョイントモデルで融合し、多種多様な意味や地理的ソースでスキーマアライメントモデルの同時学習を可能にする、新たな共有潜在空間を提案する
- 結果のモデルからtag-to-classのアライメントを抽出するアルゴリズムを開発

• 結果

- ベースラインと比較してF1スコアは37%ポイント向上した。
- 1,000万を超えるOSMのエンティティをWikidataとDBpediaのクラスにアノテーションした（現在存在するアノテーションから400%以上増加）。

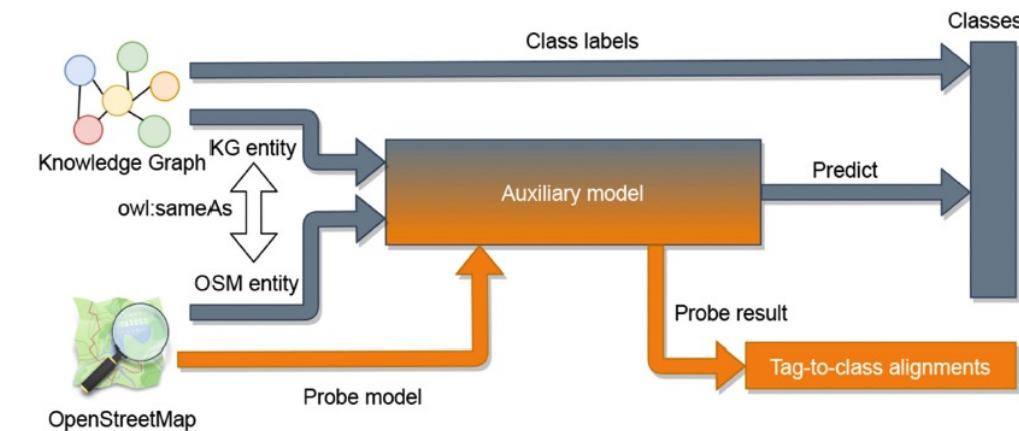


Fig. 1. Overview of the NCA architecture. The gray color indicates the first step (training of the auxiliary classification model). The orange color indicates the second step, i.e., the extraction of tag-to-class alignments. (Color figure online)

Background Knowledge in Schema Matching: Strategy vs. Data

- Jan Portisch^{1,2}, Michael Hladik³, and Heiko Paulheim¹

• ¹Data and Web Science Group University of Mannheim, Mannheim, Germany. ²SAP SE Business Technology Platform | One Domain Model, Walldorf, Germany. ³SAP SE Business Process Intelligence, Walldorf, Germany

• 一言でいうと

- スキーママッチングにおいて、戦略と背景知識のどちらが重要か、優れた戦略やリソースはどれか明らかにするためのテストベッドを提供し、実験結果を考察

• 動機

- スキーマやオントロジーのマッチングのタスクにおいて、戦略と背景知識のどちらが重要か、優れたリソースや優れた戦略があるか明らかにする。

• 手法

- 様々な背景知識のリソースと利用戦略を扱うことができる、シンプルで汎用的なマッチングプロセスを提案
- 6つの汎用的な背景知識と、3つの異なる戦略（同義語、上位概念、ナレッジグラフ埋め込み）の組み合せでマッチングタスクを評価

• 結果

- 明示的な戦略（同義語、上位概念）は潜在的な戦略（埋め込み）よりも優れていることがわかった。
- 戦略の選択は背景知識よりもアライメントに大きな影響を与えることがわかった。
- 普遍的に優れたリソースを特定することはできなかつたが、BabelNetは一貫して良い結果を出した。

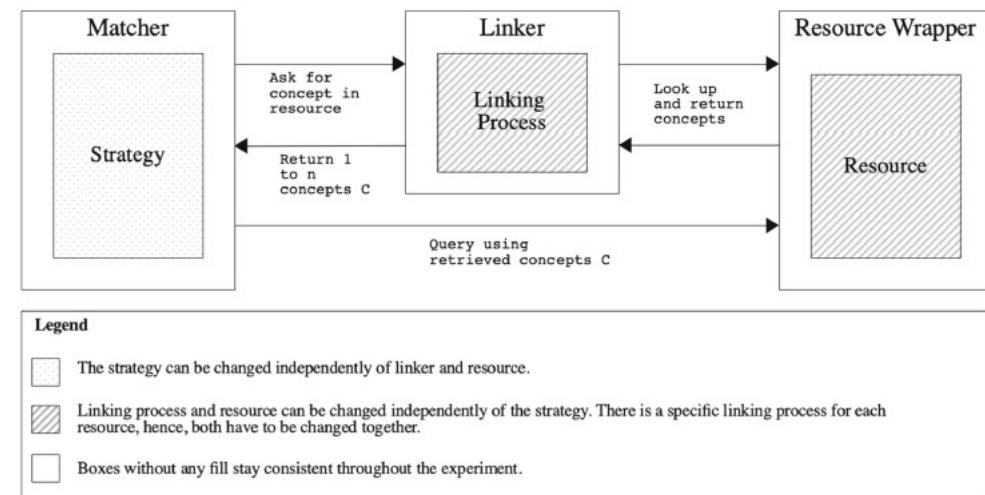


Fig. 1. Architectural setting to evaluate different background datasets exploiting different strategies.

ISWC2021サーベイ

SESSION 3B: Data Transformation

SESSION 4C: Data Analytics

藤原浩司(東芝)

SESSION 3B:

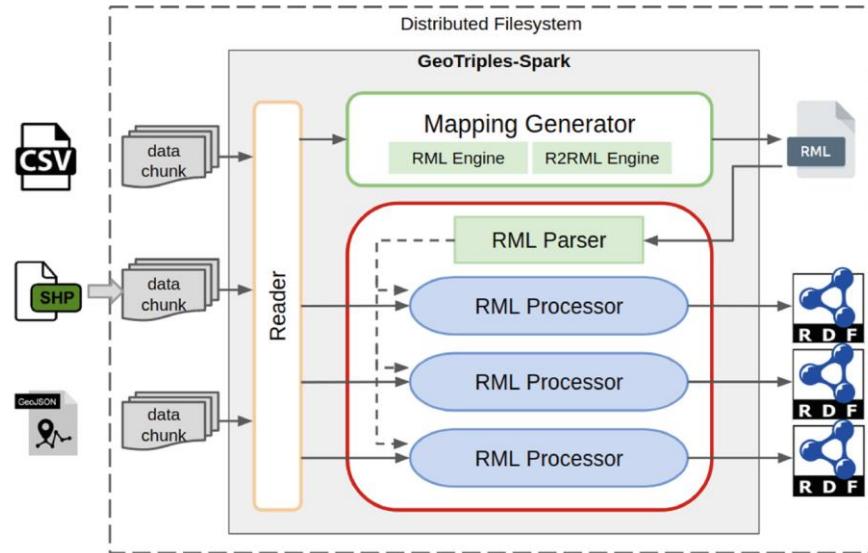
Data Transformation

A Source-to-Target Constraint rewriting for Direct Mapping (Research)

- 概要
 - Direct MappingにSQLから SHACLへの制約書き換えを追加
 - 目的
 - RDBからRDFへの変換時に従来失われていた制約情報をSHACLに変換することで保持する
 - 従来のOWLを使った変換では単調性が失われるなどの問題があった
 - 手法
 - 制約の変換をデータログによって記述し、Direct Mappingを拡張する
 - 結果
 - Direct Mappingの優位点を損なわない変換を実現
 - セマンティクスを(弱く)保持する(主キーあり、制約違反していない前提)
-
- The diagram illustrates the constraint rewriting process. It shows the flow from various input components to output formats.
- Input Components:**
 - Data constraints δ (represented by a rectangle)
 - Key constraints σ (represented by a rectangle)
 - Schema \mathcal{R} (represented by a rectangle)
 - Instance \mathcal{D} (represented by a rectangle)
 - Intermediate Components:**
 - \mathcal{M}^s (represented by a circle)
 - \mathcal{V} (represented by a square)
 - \mathcal{M}^t (represented by a circle)
 - Γ (represented by a circle)
 - OWL rules (represented by a dashed oval)
 - OWL axioms (represented by a dashed rectangle)
 - Output Components:**
 - Shapes \mathcal{S} (represented by a rectangle)
 - Graph \mathcal{G} (represented by a rectangle)
- The flow is as follows:
- Data constraints δ and Key constraints σ feed into \mathcal{M}^s .
 - \mathcal{M}^s feeds into \mathcal{V} .
 - \mathcal{V} feeds into Γ and \mathcal{M}^t .
 - Schema \mathcal{R} feeds into \mathcal{M}^t .
 - Instance \mathcal{D} feeds into \mathcal{M}^t via a dashed arrow labeled "OWL rules".
 - \mathcal{M}^t feeds into Graph \mathcal{G} .
 - Γ feeds into Shapes \mathcal{S} .
 - \mathcal{M}^t also feeds into OWL axioms, which are enclosed in a dashed rectangle.

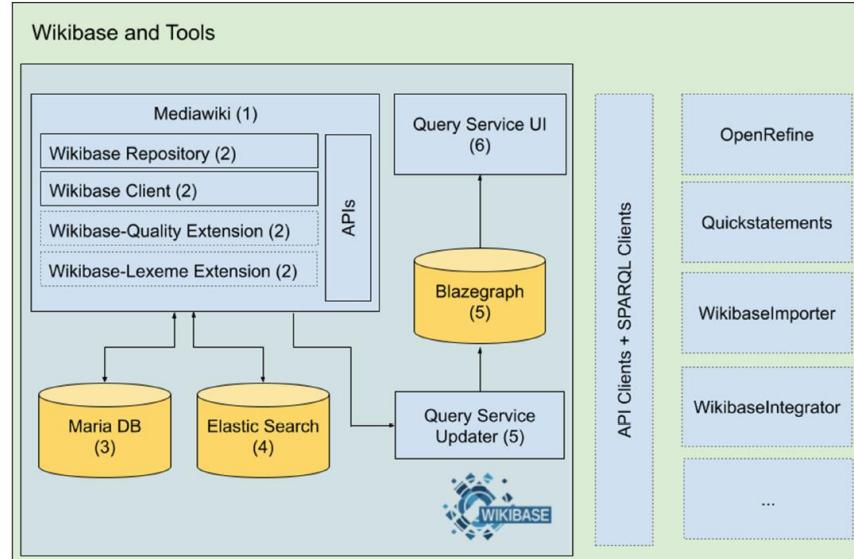
Scalable Transformation of Big Geospatial Data into Linked Data (Resource)

- 概要
 - GeoTriplesの新しいバージョンのGeoTriples-Sparkを紹介
- 目的
 - 巨大かつ更新頻度が高い地理空間データをリンクトデータ化
- 結果
 - 巨大な地理情報ファイルを妥当な時間で変換
 - CSV 1TB : 26m、ESRI 1TB : 34m
 - GeoTriples-Sparkは、衛星画像から抽出されたデータをリンクトデータに変換するためにプロジェクトExtremeEarthで使用されている



Wikibase as an Infrastructure for Knowledge Graphs: the EU Knowledge Graph (In-Use)[Best Paper Candidate]

- 概要
 - Wikidataのインフラである
Wikibaseおよび関連ツールの紹介
- 目的
 - Wikibaseを知識グラフのインフラとしての利用手法について示す
 - 特に「The EU Knowledge Graph」のプロジェクトでの利用に関して
- 結果
 - 従来のRDFインフラほど柔軟ではないが、ナレッジグラフに必要/便利で、専門家以外も使いやすい多くのサービスをすぐに提供する

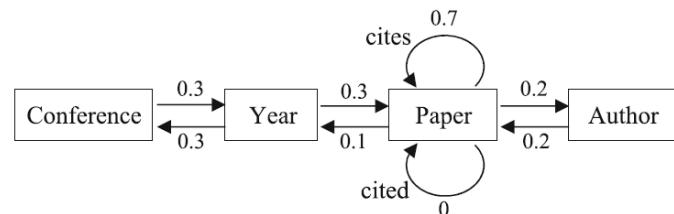


SESSION 4C:

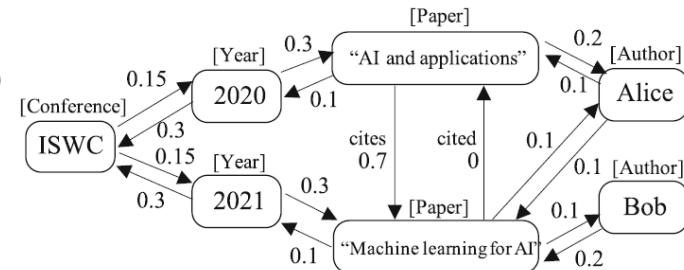
Data Analytics

Fast ObjectRank for Large Knowledge Databases (Research)

- 概要
 - 重要なノードを高速に評価するObjectRankの新しいアルゴリズムSchemaRankの提案
- 目的
 - 大規模なKGからに重要ノードを高速に得る
- 手法
 - ObjectRankによりデータを2つに分割
 - 2段階のRWR(random walk with restarts)で重要度の低いノードを段階的に削除
- 結果
 - ObjectRankの精度を犠牲にせずに、大規模なグラフに対して高速化
 - DBLPデータセットでは既存手法(FORank)よりも10倍程度はやい



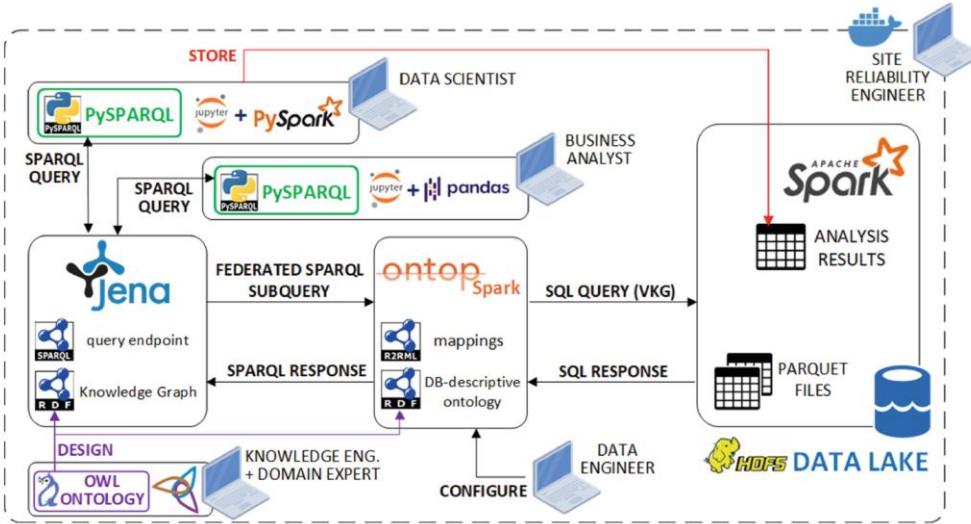
(a) Schema-graph.



(b) Data-graph

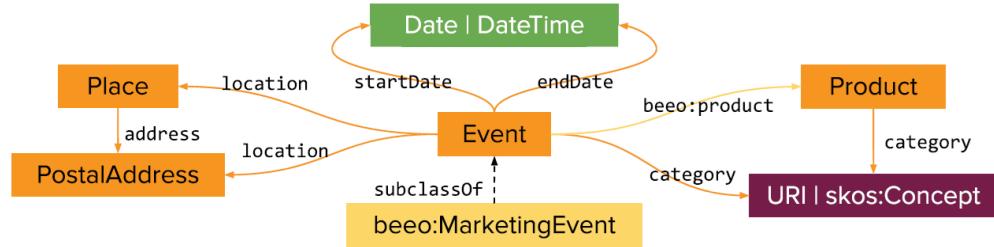
Chimera: a bridge between big data analytics and semantic technologies(Resources)[Best Paper Award]

- 概要
 - ソフトウェアスイート Chimera2を紹介
- 目的
 - ビッグデータの世界とセマンティックテクノロジーをより適切に接続する
- 結果
 - ビッグデータ解析エンジンであるSparkに格納されたKGに対して、OntopSparkを使用してSPARQLを実行できる
(従来のOntopはRDBに限定されていた)
 - PySPARQLにより、SparkDataFrame/GraphFrameの結果を得られる



BEEO: Semantic Support for Event-based Data Analytics (Resources)

- 概要
 - イベントデータオントロジー Business Event Exchange Ontology (BEOO) の紹介
- 目的
 - イベントデータの統合と分析
- 結果
 - EW-Shop(EUプロジェクト)のビジネスユースケースから作られ、様々なサービスをサポートするための共通の用語と共通のツールを提供する
 - マーケティングドメインの重要な側面をモデル化するのに効果的である
 - オープンソースとして公開されたAPIを利用、拡張することでユーザーがイベントでサービスを補強できるようになった



人工知能学会 第56回SWO研究会
ISWC2021サーベイ

2022/03/11

株式会社東芝
長野 伸一

Session 2D: Industry 2

Leveraging Knowledge Graph and DeepNER to Improve UoM Handling in Search

eBay Inc

動機

- eコマースサイトの商品検索において測定単位 (Unit of Measurement, UoM) に関するクエリ理解による検索UXの向上 (同じ長さを表す商品クエリの例 : 64.5" TV, 64 1/2 inch TV, 164cm TV)

課題

- 検索クエリのような短いテキストから測定単位の認識が困難な理由 : 複数の測定法 (メートル法とインペリアル法), 曖昧性 (gと表すグラムとギガバイト), ユーザ入力表現のばらつき

手法

- DeepNER + 知識グラフによるエンティティ解決
- KG : 商品情報とブランド, 色, 素材, UoM等を統合
- UoMサブグラフ : ラベル, 物理量, 國際単位系, 変換ルール等のエンティティ及び関係を統合

評価

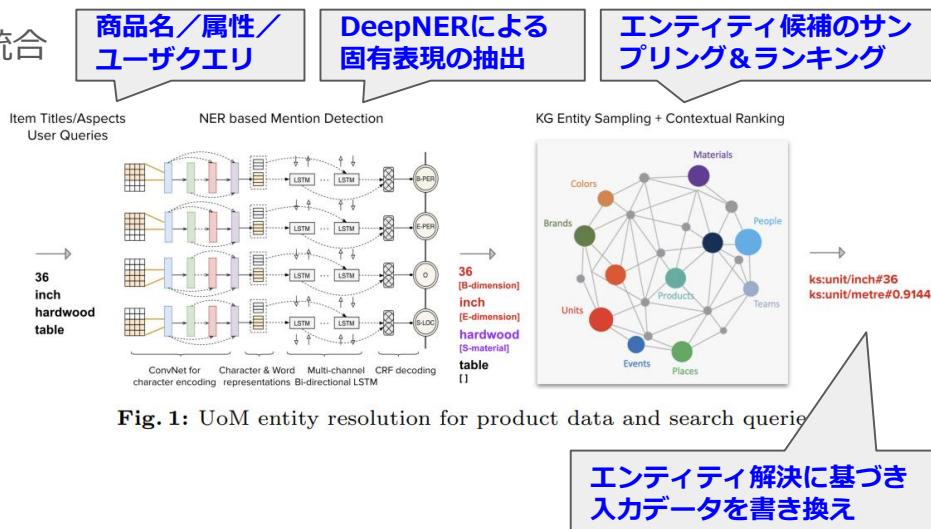
- オフライン評価で検索の再現率が大幅に向上
- オンライン評価で検索離脱率が大幅に低下

考察

- 手法の汎用性や可用性 : 議論なし

知見

- 得られた教訓・知見 : 記載なし
- 今後の課題 : サイズ, 色, 素材などへの拡張



Assessing Scientific Conferences through Knowledge Graphs

University of Cagliari, The Open University, Springer-Verlag GmbH

動機

- Springer Natureは約800件／年の会議録を発行。会議録の編集やマーケティングに必要となる国際会議の評価で生じる作業コストが課題（評価の例：会議と技術分野の関係、技術トレンド、産業界の関心等）

課題

- 既存ツール（記載なし）の機能は限定的。編集者が手作業で実施する評価作業のコスト大

手法

- Conference Dashboardを開発（ISWC2019で発表済み）
- 論文21M件、特許8M件、会議5K件の情報を統合し、Academia/Industry DynAmics Knowledge Graphを構築

評価

- Sysmte Userability Scale^{*2}のスコア87.5/100（5段階で最高）
- 編集チームによる分析作業時間が約1/2に

考察

- 手法の汎用性や可用性：議論なし

知見

- 得られた教訓・知見：記載なし
- 今後の課題：論文誌への拡張、オープン版の一般公開

*1: <http://w3id.org/aida/dashboard/>

*2: https://en.wikipedia.org/wiki/System_usability_scale



Conference Dashboard *1

対話操作可能な機能：関連会議との比較、研究トピックの年代変化、産業分野の論文比率など

Knowledge Graphs to Help with Data-driven Clinical Decision-making

Accenture Labs, Accenture Applied Intelligence

動機

- 医者が患者の治療法を選択する際の意思決定を支援するために、過去の類似症例を分類する

課題

- 既存技術における課題の記載なし

手法

- KGを構築
 - ノード：患者(17,948), 処方薬(17,400), 病状(38,930), 医療処置(279)
 - リンク：hasPrescription(125万), hasDiagnosis(153万), hasProdecure(36128), 他(50万)
- AmpliGraphで200次元のベクトルに埋込

評価

- クラスタリング結果の定性的評価どまり

考察

- 手法の汎用性や可用性：議論なし

知見

- 得られた教訓・知見：記載なし
- 今後の課題：記載なし

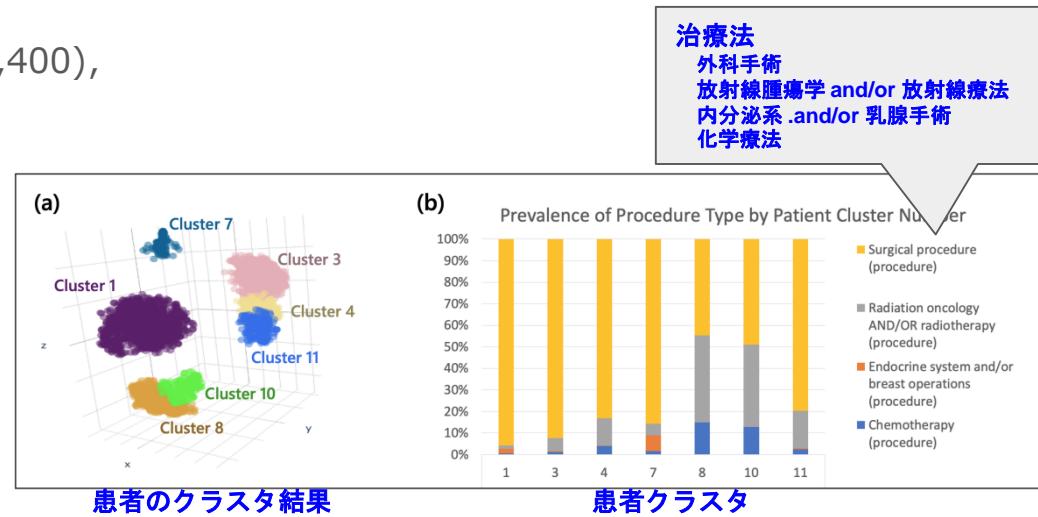


Fig. 1. (a) A subset of the patient clusters generated after projecting the embeddings into a 3D space. (b) The prevalence of different types of procedures (therapies) in patient medical records varied strongly by cluster.

A Healthcare Knowledge Graph-based Approach to Enable Focused Clinical Search

Elsevier Health and Commercial Markets

動機

- 臨床医は患者の診断や治療において、焦点は定まっているものの難解な疑問に対して、幅広い文献ソースから情報を探し、それらを合成する。臨床医による検索行為を支援し、正確で最新かつ信頼性の高い情報を的確に絞り込む検索エンジンを提供する
- Focused Searchの例：“Drug for condition X?”, “Cause of symptom Y?”

課題

- 既存技術における課題の記載なし

手法

- Elsevier Healthcare Knowledge Graphを利用した臨床情報向け Focused Searchサービスを改良
- KG：医療概念(40万), 医療用語(150万), 関係(800万)
- 専門家/NLPによる最新文献へのタグ付け, パイプラインによるKGの更新

評価

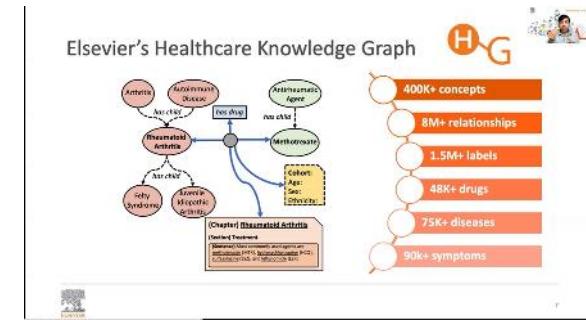
- ランキング評価指標nDCGの向上を確認
- Elsevierの医療情報検索サイトClinicalKeyのFocused Searchに搭載済み

考察

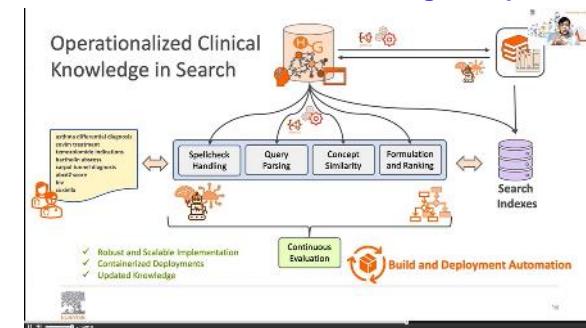
- ベクトル埋込みを利用することにより意味レベルの検索が可能に

知見

- 得られた教訓・知見：パイプライン化による早期更新. “manual curation”は欠かせない
- 今後の課題：記載なし



Elsevier Healthcare Knowledge Graph *1



Focused Searchのプロセス *1

*1: http://videolectures.net/iswc2021_2d_clinical_search/

Session 3D: Industry 3

BAMM Aspect Meta Model

Robert Bosch GmbH, Bosch.IO GmbH

動機

- 製造分野のデジタルツインにおける物理資産（例：掘削機）と抽象資産（例：生産工程）の相互運用性を実現する

課題

- 産業データ辞書の標準規格であるIEC 61360, IEC CDD, ECLASSは、文脈におけるデータの意味情報を解釈する問題を解決していない

手法

- 資産に対するビューを定義するアスペクトのメタモデルを提案
- アスペクトモデルには下記を記述
 - 実行時データの構造に関する情報（例：データには「temperature」というプロパティがあり、それが数値を持つ）
 - 実行時データに含まれない情報（例：物理単位や値の範囲）
- メタモデルはSHACLで記述、アスペクトモデルはRDF/Turtleで記述

評価

- 独Bosch社内で実証：50万種、1,500万件の製品のアスペクトをモデル化
- Open Manufacturing Platform^{*1}でメタモデルとSDKを公開

考察

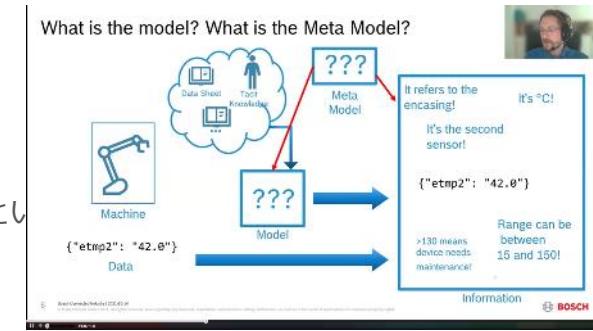
- 手法の汎用性や可用性：議論なし

知見

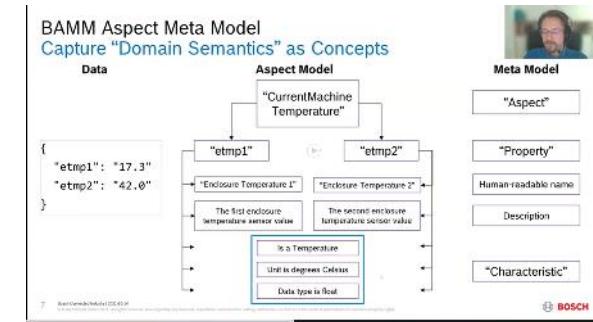
- 得られた教訓・知見：記載なし
- 今後の課題：Bosch Semantic Stackに組み込み、工場DigitalTwinへ応用

*1: <https://open-manufacturing.org/>

*2: http://videolectures.net/iswc2021_3d_meta_model/



モデルとメタモデル²



アスペクトモデル²

OntoScene: Ontology Guided Indoor Scene Understanding for Cognitive Robotic Tasks

TCS Research

動機

- ロボットが人間との共存空間でタスク実行するために信頼性、解釈性の高いシーン理解を実現する

課題

- RGB-Dカメラなどのセンサ入力のみで世界（シーンの連続）を意味的に理解することは困難
- End-to-endの機械学習アプローチはシンボリック手法よりも解釈性で劣る

手法

- OntoSceneオントロジを開発
- ベース：室内環境向け認知タスクに特化した抽象オントロジCORA
- 追加拡張した知識の例：物体の相対位置、移動性、属性と認識アルゴリズムのリンクなど

評価

- 評価や実績に関する記載なし

考察

- 機械学習に対する優位性を定性的に述べているが、オントロジ構築の汎用性・可用性に対する具体的な議論なし

知見

- 得られた教訓・知見：記載なし
- 今後の課題：記載なし

ユーザ指示「ボトルの左にあるカップを探す」に対し、まずオブジェクト検出アルゴリズムを呼び出して「カップ」を検出し、次にオントロジを参照してプロパティのリストを取得し、シンググラフを生成する。

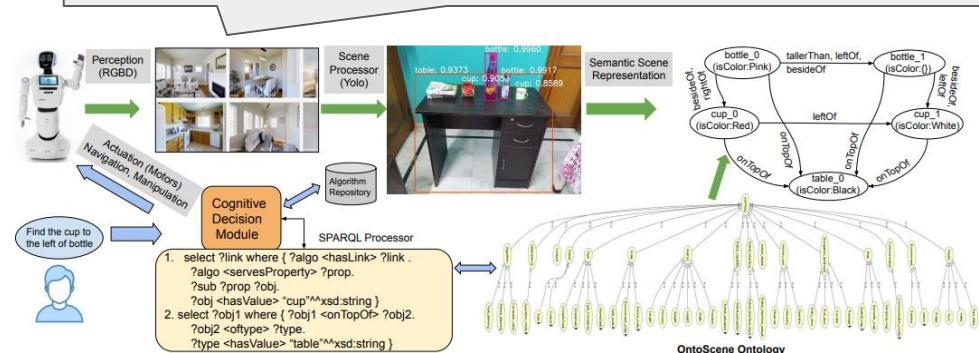


Fig. 1. System using OntoScene for scene understanding to aid in robotic tasks

Enabling a Semantic Sensor Knowledge Approach for Quality Control Support in Cleanrooms

University of Salerno, Leonardo SpA

動機

- クリーンルームの品質管理を対象に、製品レベルの品質管理を必要とする環境下でセンサデータのリアルタイムな収集・分析にセマンティック技術の活用が有用であることを実証する

課題

- センサデータの格納に対する要件：
 - 後で品質管理に必要なデータを取り出せること
 - 直近および過去の時間を参照できること

手法

- 著者らが[3]で提案のcontextual sensor network ontologyを拡張し、センサデータの集合を表すクラスを追加

評価

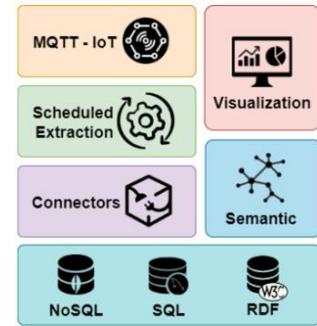
- 伊Leonardo SpA社のクリーンルームで実証

考察

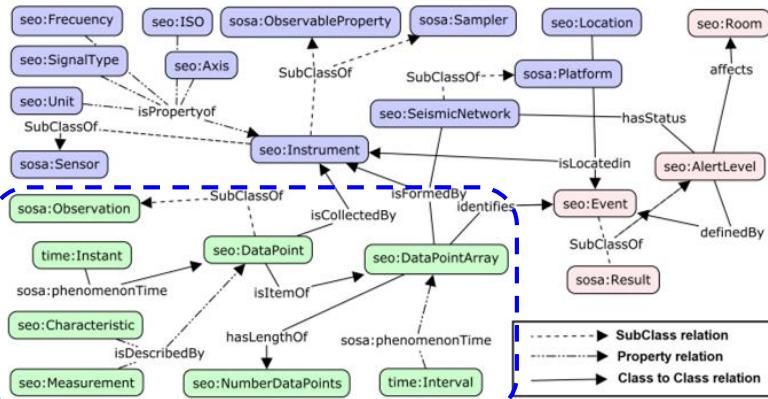
- レポート生成や予測分析での活用、および製品情報や作業員の行動情報との連携について言及しているが、手法の汎用性や可用性に対する具体的な議論なし

知見

- 得られた教訓・知見：記載なし
- 今後の課題：記載なし



(a) High Level Architecture view



(b) Knowledge-Base Schema

A Framework for Automatically Interpreting Tabular Data at Orange

Orange, EUROCOM

動機

- 企業で扱われる多様な形式の表データを知識グラフとマッチングさせ、検索/推薦/NLPなど多様な知識処理に活用する

課題

- 既存技術における課題の記載なし
- 多様な表形式/言語/用語が混在し、用途に適した表データの特定が困難

手法

- 表データにアノテーションするDAGOBAHを開発
- 表が持つ文脈、及び知識グラフが持つ文脈を埋込みベクトル空間上で比較することで曖昧性を解消

評価

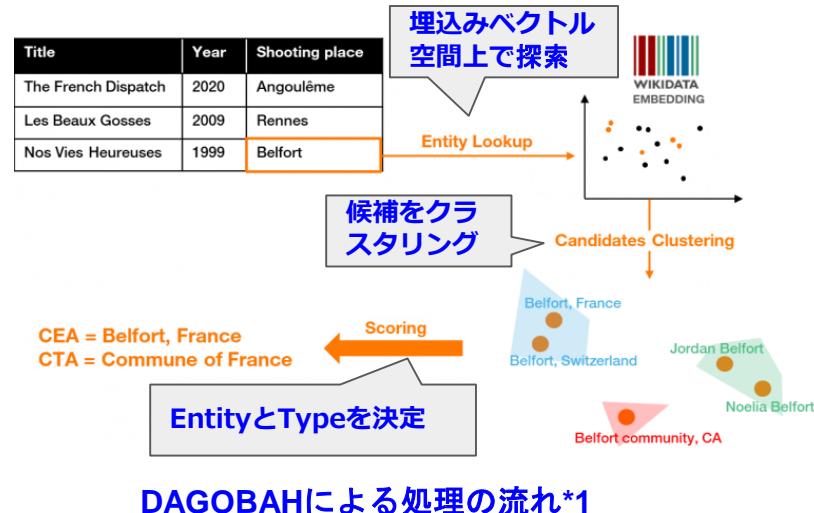
- 研究コンペSemTab2019,2020で競争力を証明

考察

- 均質できれいな表データ、およびWikiData等のオープンなKGの活用を前提とし、適用は限定的

知見

- 得られた教訓・知見：記載なし
- 今後の課題：企業等が扱う実データへの適用やドメイン知識の活用



*1: <https://hellofuture.orange.com/en/dagobah-make-tabular-data-speak-great-again/>

ISWC2021サーバイ

SESSION 1D: Industry1

SESSION 4A: Ontologies

砂川英一(東芝)

SESSION 1D:

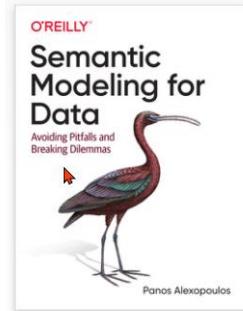
Industry1

Building a Semantic Modeling Team - Skills to Have and Attitudes to Avoid (Industry)

Proc.に原稿なし

KGに関わるメンバーは概念とデータの両方を扱えるべし

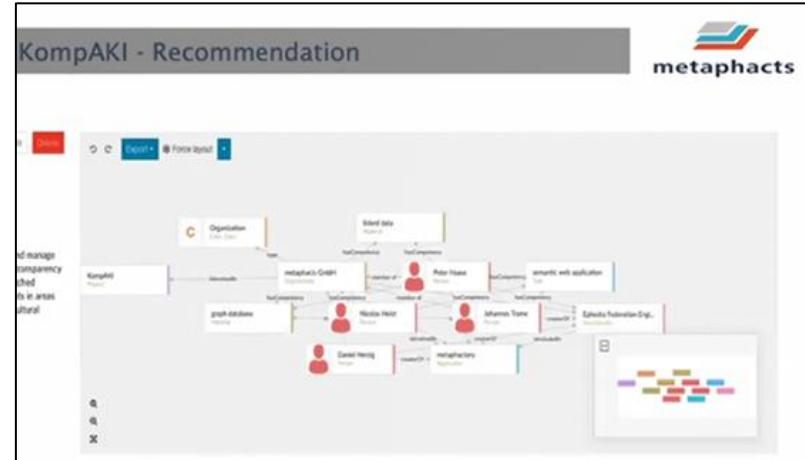
- 動機
 - オントロジー屋, 論理屋, 統計データ屋・・・
Semantic Modelには様々な人が関わる
 - その中で, 正確, 明示的, 共通理解となるモデルを作るのは難しい
- 課題
 - チームとして持つべきスキルや姿勢が不明
- 知見
 - 概念・意味思考のスキル=モデリング言語を正しく理解し, 重要な意味の明示化する
 - データ工学スキル=情報抽出, NLP, 機械学習など.
非専門家がモデルと会話するためUX/UIデザインも必要
 - 避けるべき姿勢：
学者ぶって細部にこだわりすぎ, 逆に大事な意味の差を考えない,
全て真理値で表現できると思う, データ・専門家知識・集合知への偏り思考,
全てを手元の技術で解決したがる



Flexible and Extensible Competency Management with Knowledge Graphs (Industry)

KGを使って、AI関連の技術者マップ／Know-whoを作った

- 動機
 - AI分野で必要な能力を明確化し、組織に広める
- 課題
 - 現在は、特定のトピックに詳しい専門家を個人の繋がりを辿って見つけてくるしかない
- 手法
 - AI分野の人や組織構造が持つ能力を、オントロジー（Competency KG）を使って表現
 - metaphactoryベースのアプリで利用
 - ビジネス課題から関連能力を特定、特定の能力を持つ人や組織を見つける、など
- 考察
 - Metaphactoryを使ったので、外部ソース（Microsoft Academic Graph）との連携や、他ドメインへの拡張も容易である
 - (competency managementについては不明)



Boosting Information Extraction through Semantic Technologies: The KIDs use case at CONSOB (Industry)

イタリアにおける
公正取引委員会みたいなもの

文書からの情報抽出の処理高速化・ルール記述容易化をKGで実現

- 動機
 - 100万件／年の金融商品に関するPDF文書（KIDs : Key Information Documents）から、規定の情報を抽出するのが大変（部分的な自動化）
- 課題
 - 質の高い情報を抽出するための手作業コストが高い
- 手法
 - ドメイン専門家と設計したKGに、ルールベースで自動抽出した要素を流し込む
- 評価, 考察
 - 12の項目（商品名, 発行日など）で抽出を試行
 - CoreNLP→TokenRegex→KGのFacts化したら網羅性・再現性は99%で良かったが、実行時間とルール構築が問題
 - MASTRO SYSTEM-T（KGによる抽出ツール）を使って両課題をクリア
 - 言語に対するルールの記述性、それを抽出処理のアサーションに変換できる

Integrating Graph and Machine Learning for Fraud Detection Use Case

KGとMLを融合させ決済詐欺の防止に適用

- 動機
 - 人や業者のアカウントと取引から、決済詐欺を検知・防止するPFを構築する
- 課題
 - 従来MLでは表データに依存（＝固定的）
 - データの時間変化
 - グラフDBのデータサイエンティストにとっての使いやすさ
- 手法
 - ユーザ・業者のアカウント情報と間の関係をグラフデータ化
 - アカウントを結びつける経路をオンデマンドで増減できる
 - 理解しやすく、MLアルゴリズムの向上にも寄与する
 - グラフに時間属性を持たせ、最新スナップショット（Graph）と過去履歴（RDB）を結ぶ
 - GraphQLをサービスAPIとして採用
- 評価
 - 適合率・再現率ともに50%向上させた
 - グラフから得られた特徴量が、詐欺検出で2番目に重要な特徴量だった

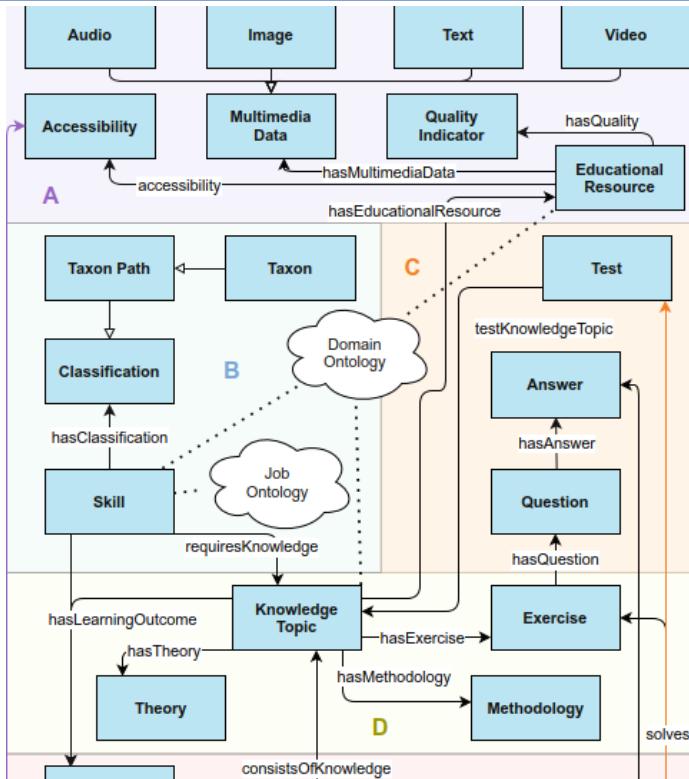
SESSION 4A:

Ontologies

EduCOR: An Educational and Career-Oriented Recommendation Ontology (Resources)

概念定義にパターンがあるオントロジーで、個人に合わせた教育コンテンツをリコメンド

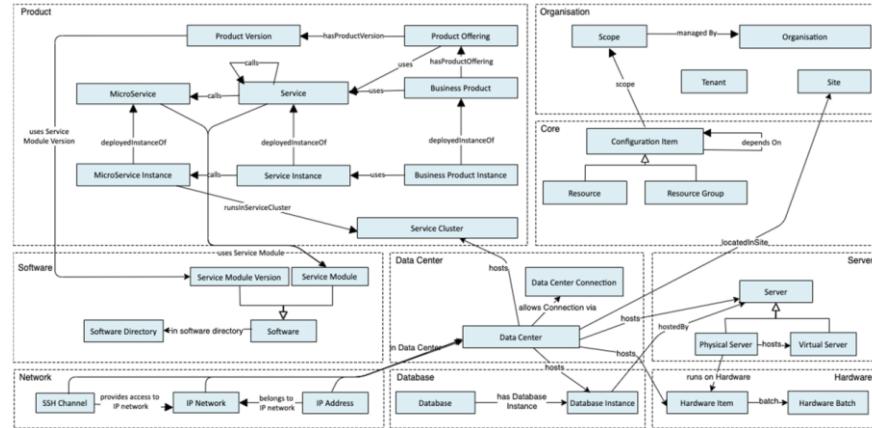
- 動機
 - eラーニングのPF・リポジトリを横断し、個人に合わせた学習リソースを推薦する
 - Open Educational Resources(OER)やPF化は進んだが、オントロジーやスキーマが欠落
- リソースの設計方針
 - SemanticWebコミュニティがこれまで開発した幅広いドメインのオントロジーを学習リソースに繋ぐ
- 新規性
 - 教育のためのオープンフリーなオントロジーである
 - コミュニティ標準となるオントロジーのコンポーネント（定義のパターン）を提供
- 品質
 - 複数のOERに対して高いカバー率を達成できた



A High-Level Ontology Network for ICT Infrastructures (Resources)

ICTインフラを包括的に扱うオントロジーを事業データスキーマから構築

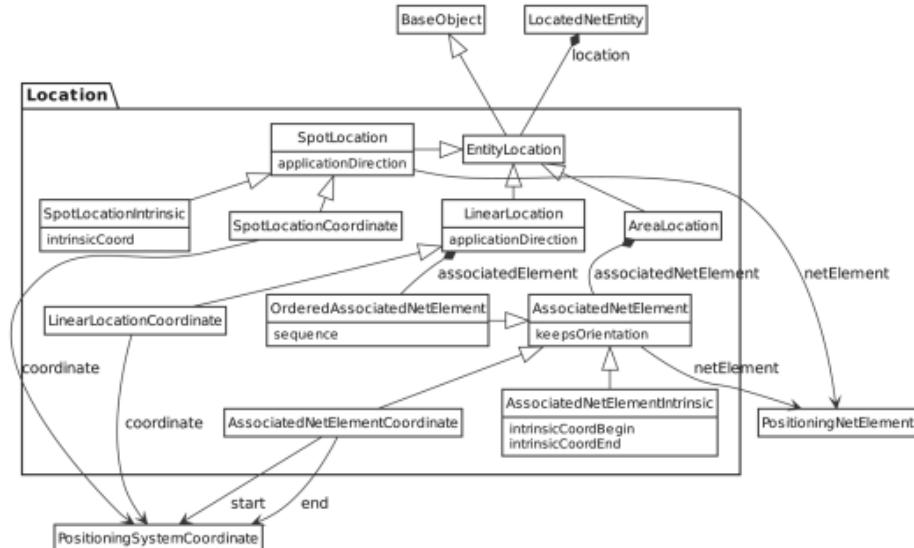
- 動機
 - ICTサービスのインフラリソース(HW/SW)をカタログ化→構成管理を容易にしたい
- リソースの設計方針
 - Linked Open Terms (LOT) の手法・技術を用い、要件定義→実装→公開→メンテ
 - Huaweiの既存情報（インフラDB、共通CSV）をもとに作成
- 新規性
 - ICTインフラに広く使える共通のエンティティと関係をオントロジー化できた
- リソースの再利用性/活用実績、可用性
 - Huaweiのクラウド管理、DevOps基盤で活用

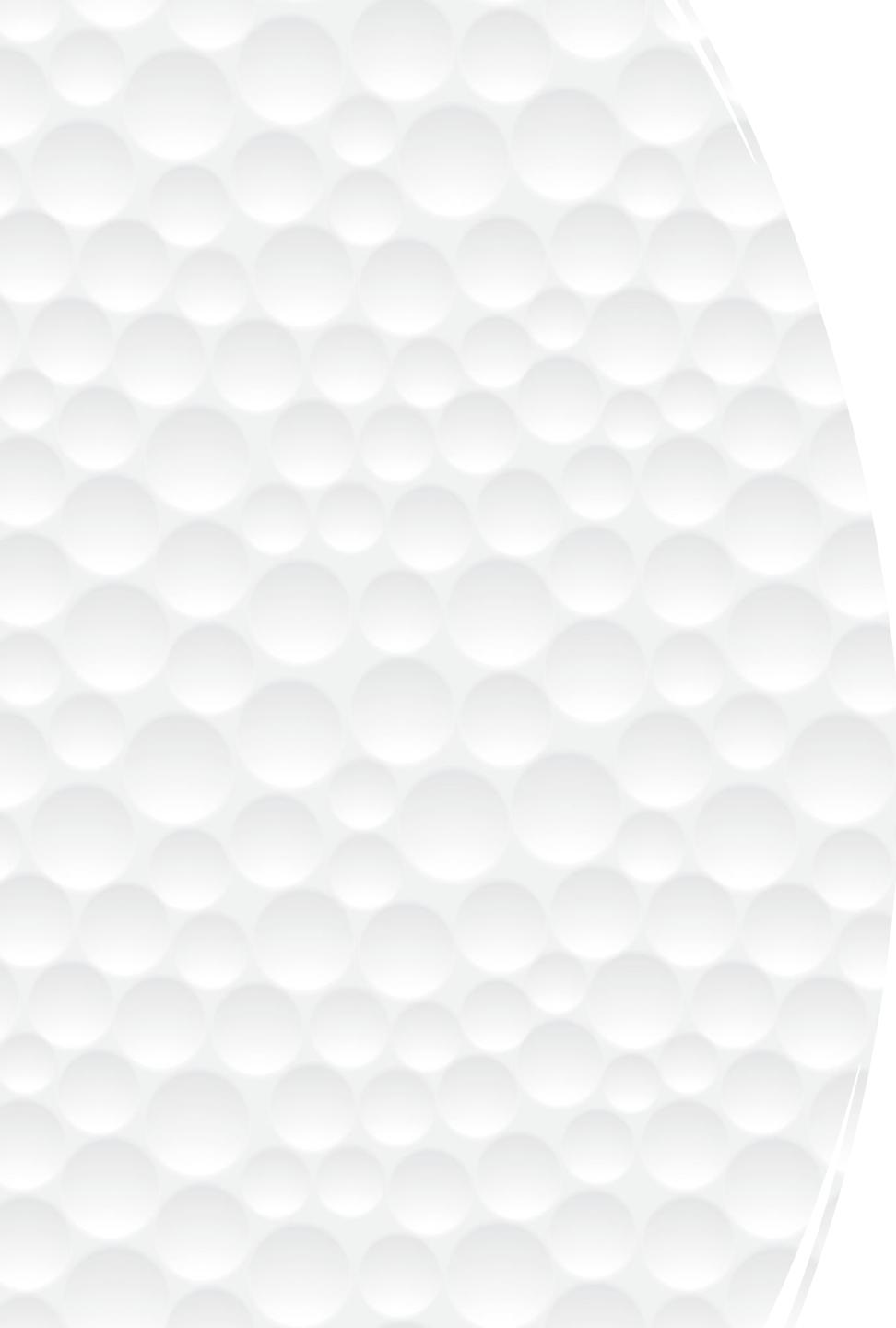


Rail Topology Ontology: A Rail Infrastructure Base Ontology (Resources)

鉄道分野のデータ交換に関する国際規格モデルのオントロジー化

- 動機
 - 鉄道のデータ表現を一貫させ、エンジニアリングツール開発コストを低減
- リソースの設計方針
 - 経路に関するcompetency question
 - 鉄道網トポロジー、インフラ要素、集約関係を含む
 - 既存の鉄道データ交換の国際標準 (IRIS30100)に準拠
 - このUMLモデルをもとに作成
- 新規性
 - 設計方針を満たすものが無い
- リソースの再利用性/活用実績、可用性
 - 欧州鉄道機関ERAが所管する基本レジストリを統合
 - 参考にした規格に依存する他の鉄道関連モデルのオントロジー化にも寄与





ISWC2021

サーべイ

Session 5A: Linked Data Applications
Session 5B: Validation

山本 泰智(ROIS-DS-DBCLS)
@yayamamo

The Punya Platform: Building Mobile Research Apps with Linked Data and Semantic Features

- Evan Patton, William Van Woensel, Oshani Seneviratne, Giuseppe Loseto, Floriano Scioscia and Lalana Kagal
- Resources
- Session 5A: Linked Data Applications
- Best Paper Candidate
- <https://github.com/knowledgegraph-yra/Survey/issues/4>

The Punya Platform: Building Mobile Research Apps with Linked Data and Semantic Features

- ・スマホのアプリを簡単に開発できる環境Punyaを紹介。
- ・オープンソースのウェブアプリで、Linked Dataを利用しやすいようにGraphQLやSPARQLをサポートしたり、BLEを利用してIoTに対応しやすくしたりしている。
- ・生命科学研究者など、開発エンジニアではない利用者を想定。
- ・<http://punya.mit.edu/> で利用可能。
- ・現在はAndroid OSのみだが、iOSにも対応予定。

Reconciling and Using Historical Person Registers as Linked Open Data in the AcademySampo Portal and Data Service

- Petri Leskinen and Eero Hyvönen
- In-Use
- Session 5A: Linked Data Applications
- <https://github.com/knowledgegraph-yra/Survey/issues/8>

Reconciling and Using Historical Person Registers as Linked Open Data in the AcademySampo Portal and Data Service

- ・フィンランドの学術界で活躍した人々の系譜を、個々人の登録情報から自動的に抽出して生成。
- ・過去に学術界で研究活動していた人々の系譜をLinked Dataとして保存し、デジタル人文学データとしてポータルサイトで情報公開することが目的。
- ・親族に関する情報を利用することで各人に関する情報の曖昧性（個人名を特定しない呼称など）を解消。
- ・ポータルサイト <https://www.ldf.fi/dataset/yoma>
- ・SPARQLエンドポイント <http://ldf.fi/yoma/sparql>

Mapping Manuscript Migrations on the Semantic Web: A Semantic Portal and Linked Open Data Service for Pre-modern Manuscript Research

- Eero Hyvönen, Esko Ikkala, Mikko Koho, Jouni Tuominen, Toby Burrows, Lynn Ransom and Hanno Wijsman
- In-Use
- Session 5A: Linked Data Applications
- Best Paper Candidate
- <https://github.com/knowledgegraph-yra/Survey/issues/9>

Mapping Manuscript Migrations on the Semantic Web: A Semantic Portal and Linked Open Data Service for Pre-modern Manuscript Research

- ・近代以前に発表された写本の情報をリンクトデータ化し、ポータルサイトで検索可能とした。
- ・さまざまな観点から絞り込めるファセット検索機能が充実。
- ・最近デジタル化されている写本のメタデータは形式がばらばらで、メタデータ横断的に調査することが難しかった。
- ・ポータルサイト
<https://mappingmanuscriptmigrations.org/en/>
- ・SPARQLエンドポイント <http://ldf.fi/mmm/sparql>

Dataset or Not? A study on the veracity of semantic markup for dataset pages

- Tarfah Alrashed, Dimitris Paparas, Omar Benjelloun, Ying Sheng and Natasha Noy
- Research
- Session 5B: Validation
- Best Student Paper Award
- <https://github.com/knowledgegraph-yra/Survey/issues/5>

Dataset or Not? A study on the veracity of semantic markup for dataset pages

- Schema.org の Schema.org/Dataset タグが付けられているWebページについて、実際にはデータセットではない例が散見されるので、それを見分ける分類器を開発した。
- その結果、AUC -PRで0.99と、非常に高い精度で分類できた。
- Google Dataset Searchの検索対象を選定する際に利用を想定。
- Webページの記載内容を分類する手法が多く提案されているが、データセットであるか否かを分類する試みは初めて。

A Framework for Quality Assessment of Semantic Annotations of Tabular Data

- Roberto Avogadro, Marco Cremaschi, Ernesto Jimenez-Ruiz and Anisa Rula
- Resources (Software Framework)
- Session 5B: Validation
- <https://github.com/knowledgegraph-yra/Survey/issues/6>

A Framework for Quality Assessment of Semantic Annotations of Tabular Data

- 多数の表形式のデータが公開されており、そこから意味的な情報を抽出して、より構造化されたデータにする試みは多い。
- しかし、抽出処理に着目した評価法は検討されていなかった。
- そこで、構造化の程度や質について評価する方法とツール STILTool 2を提案した。
- ツールの入手先 https://bitbucket.org/disco_unimib/stiltool
- Dockerイメージ <https://hub.docker.com/r/cremarco/stiltool>

ProGS: Property Graph Shapes Language

- Philipp Seifer, Ralf Lämmel and Steffen Staab
- Research
- Session 5B: Validation
- <https://github.com/knowledgegraph-yra/Survey/issues/7>

ProGS: Property Graph Shapes Language

- RDFデータの検証を行うためにSHACLやShExがあるが、プロパティグラフにおいては同等の言語がない。
- そこで、 SHACLを基にしてProGSという言語を設計した。
- 実装の計算量はNP完全であるため、 SHACLのRDF検証器と同等であり、さらに、表現力は高まっている。
- 動作確認用の実装はオープンソースで公開
<https://github.com/softlang/progs>

ISWC2021サーベイセッション

Session 1A

Session 2B

2022年3月11日

鵜飼孝典(富士通株式会社)

Session 1A – Question Answering



- Open Domain Question Answering over Knowledge Graphs using Keyword Search, Answer Type Prediction, SPARQL and Pre-trained Neural Models (Reserch)
 - Christos Nikas, Pavlos Fafalios and Yannis Tzitzikas.
- EDG-based Question Decomposition for Complex Question Answering over Knowledge Bases (Research)
 - Xixin Hu, Yiheng Shu, Xiang Huang and Yuzhong Qu.
- Generative Relation Linking for Question Answering over Knowledge Bases (Research)
 - Gaetano Rossiello, Nandana Mihindukulasooriya, Ibrahim Abdelaziz, Mihaela Bornea, Alfio Massimiliano Gliozzo, Tahira Naseem and Pavan Kapanipathi.

■ 著者/所属機関

- Christos Nikas(1,2), Pavlos Fafalios(1) and Yannis Tzitzikas(1,2)
 1. Information Systems Laboratory, FORTH-ICS, Heraklion, Greece
 2. Computer Science Department, University of Crete, Heraklion, Greece

■ 背景

- 知識グラフに対するQAは、エンドユーザーにとって適切で、満足度の高い、喜ばれるものになりにくい

■ 手法

- RDF上の汎用的な対話型キーワード検索システムを補完するパイプライン
- 1)RDF上の汎用エンティティ検索サービス、2)回答のタイプ予測、3)SPARQLによるエンリッチメント、4)事前学習済みニューラルモデル

■ ソースコード、デモ

- ソースコード : <https://github.com/cnikas/isl-smart-task>
- デモ : <https://demos.isl.ics.forth.gr/elas4rdf/>

■ 結果

- Precision by 6%, Recall by 7% and F1 score by 7% 向上 (over WebQuestions)
- Entity search tasks by 6% 向上 NDCG@100 (over DBpedia Entity dataset).

EDG-based Question Decomposition for Complex Question Answering over Knowledge Bases (Research)

FUJITSU

■ 一言でいうと

- 論文内容を一言で要約

■ 著者/所属機関

- Xixin Hu, Yiheng Shu, Xiang Huang and Yuzhong Qu
 - State Key Laboratory for Novel Software Technology Nanjing University Nanjing, China

■ 動機

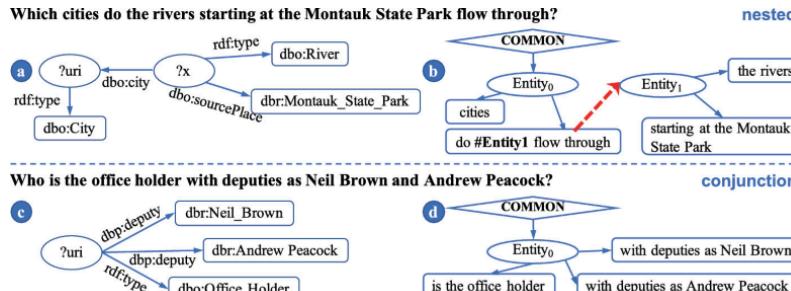
- 複数のKBの関係や制約を必要とする複雑な質問に対して、知識ベース質問応答は質問の理解、構成要素のリンク(例:実体、関係、型のリンク)、クエリの合成など多くの課題がある

■ 手法

- 複雑な質問の構造を表現するために、Entity Description Graph (EDG) と呼ばれる新しいグラフ構造を提案し、EDG構造を利用して、EDGQAと呼ばれるDBpedia上のQAシステムを実装する

■ ソース

- <https://github.com/HXX97/EDGQA>



出典 https://link.springer.com/chapter/10.1007/978-3-030-88361-4_8 Fig1

Table 2.

QA performance on LC-QuAD and QALD-9

Approaches	LC-QuAD 1.0			QALD-9			
	P	R	F	P	R	F	F-Q
WDAqua [7]	0.22	0.38	0.28	0.261	0.267	0.250	0.289
QAmp [24]	0.25	0.50	0.33	—	—	—	—
gAnswer [9]	—	—	—	0.293	0.327	0.298	0.430
NSQA [11]	0.38	0.40	0.38	0.314	0.322	0.309	0.453
Pointer network* [20]	0.220	0.241	0.230	0.096	0.109	0.102	0.190
EDGQA	0.505	0.560	0.531	0.313	0.403	0.320	0.461

FQ: Macro-F1-QALD measure, *: our implementation

Generative Relation Linking for Question Answering over Knowledge Bases (Research)

FUJITSU

■ 一言でいうと

- 論文内容を一言で要約

■ 著者/所属機関

- Gaetano Rossiello, Nandana Mihindukulasooriya, Ibrahim Abdelaziz, Mihaela Bornea, Alfio Massimiliano Gliozzo, Tahira Naseem and Pavan Kapanipathi
 - IBM Research, T.J. Watson Research Center, Yorktown Heights, NY, USA

■ 動機

- 知識ベースに対する質問応答を可能にするためには、関係性のリンクが不可欠

■ 手法

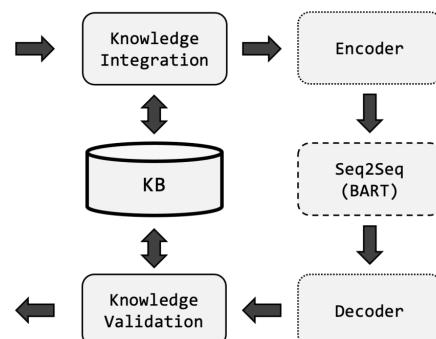
- 関係リンクを生成問題として捉え、事前に学習されたシーケンス間モデルを用いることで、関係リンクのための新しいアプローチを提案

■ ソース

- <https://github.com/IBM/kbqa-relation-linking>

Input: Question

What is the owning organization
of the Ford Kansas City Assembly
Plant and also the builder of the
Ford Y-block engine?



	LC-QuAD 1.0			QALD-9		
	P	R	F1	P	R	F1
Falcon 1.0 Sakor et al. [2019]	0.42	0.44	0.43 ⁹	0.23	0.23	0.23 ⁹
SLING Mihindukulasooriya et al. [2020]	0.41	0.55	0.47 ⁹	0.39	0.50	0.44 ⁹
GenRL wo/KB	0.47	0.50	0.48	0.51	0.43	0.47
GenRL	0.54	0.74	0.60	0.49	0.61	0.53

Figure 2: GenRL Framework

Session 2B – Learning from Wikidata



■ Learning to Predict the Departure Dynamics of Wikidata Editors (Research)

- Guangyuan Piao and Weipeng Huang

■ (Best Student Paper Candidate) Learning to Recommend Items to Wikidata Editors (Research)

- Kholoud Alghamdi, Miaojing Shi and Elena Simperl

■ (Best Paper Award) Improving Inductive Link Prediction Using Hyper-Relational Facts (Research)

- Mehdi Ali, Max Berrendorf, Mikhail Galkin, Veronika Thost, Tengfei Ma, Volker Tresp and Jens Lehmann

■ 著者/所属機関

- Guangyuan Piao(1), Weipeng Huang(2)
 - Department of Computer Science Maynooth University Maynooth Ireland
 - Insight Centre for Data Analytics University College Dublin Dublin Ireland

■ 動機

- Wikidataは多数のボランティア編集者からなるコミュニティによって共同開発・維持されているため、編集者の離脱を理解し予測することは非常に重要

■ 手法

- Wikidata編集者がプラットフォームに留まるか離れるかを予測するために、DeepFMを分類モデルにたいして、統計的特徴と、パターンベースの特徴という2つの異なるタイプの特徴の相乗効果を利用する
 - DeepFM : Factorization Machines (FM) と Deep Neural Networks (DNN)を組み合わせたモデル.
 - 統計的特徴: Wikidataの編集者の編集履歴を利用する
 - パターンベースの特徴: 文献[1]に示された編集をやめる人の編集履歴のパターンを利用する

Method	AUROC	F1
GBT-Zh	0.8890	0.8205
kNN-Zh	0.8731	0.7935
RF-Sa	0.7647	0.7769
LR-Sa	0.7656	0.7795
SVM-Ar	0.8396	0.8029
DeepFM-Pattern	0.8786 ± 0.0002	0.7992 ± 0.0028
DeepFM-Stat	0.8928 ± 0.0001	0.8247 ± 0.0006
DeepFM-Stat+Pattern	0.9561 ± 0.0005	0.8843 ± 0.0012

ソース

<https://github.com/parklize/ISWC2021-Learning-to-Predict-the-Departure-Dynamics-of-Wikidata-Editors>

[1] Arelli, H., Spezzano, F.: Who will stop contributing? predicting inactive editors in Wikipedia.

In: 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 355–358. IEEE (2017)

- 著者/所属機関
 - Kholoud Alghamdi, Miaojing Shi and Elena Simperl
 - King's College London, London, UK
- 動機
 - ウィキデータ項目に対する推薦システムWikidataRecを提案
- 手法
 - コンテンツベースと協調フィルタリングのハイブリッド技術を用い、アイテムの特徴とアイテム・編集者の過去のやりとりに基づき、編集者のためにアイテムをランク付けする
 - ニューラル・ミクスチャーと名付けられたニューラルネットワークは、アイテムベースの表現のための重みを学習し、アイテムと編集者の相互作用にしたがって最適化するように設計されている
- 結果
 - 2つのベンチマークデータセット
 - 220,000人の編集者が1400万件のアイテムとのインタラクションを含む汎用的なもの
 - アクティブな8000人以上の編集者の貢献度に焦点を当てたもの

Precision@k and Recall@k results comparison between our model and state-of-the-art for *Wikidata-14M*.

	Precision @k		Recall @k		
	5	10	50	100	200
GMF	0.096	0.071	0.135	0.183	0.274
BPR-MF	0.050	0.043	0.093	0.135	0.190
eALS	0.048	0.027	0.087	0.112	0.154
YouTube-DNN	0.105	0.082	0.142	0.204	0.305
WikidataRec	0.120	0.113	0.215	0.243	0.337

ソース https://github.com/WikidataRec-developer/Wikidata_Recommender

(Best Paper Award) Improving Inductive Link Prediction Using Hyper-Relational Facts (Research)



■ 著者/所属機関

- Mehdi Ali(1,2), Max Berrendorf(3), Mikhail Galkin(4), Veronika Thost(5), Tengfei Ma(5), Volker Tresp(3,6) and Jens Lehmann(1,2)
 - 1) Smart Data Analytics Group, University of Bonn, Germany 2)Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), Sankt Augustin and Dresden, Germany 3) Ludwig-Maximilians-Universität München, Munich, Germany 4) Mila, McGill University 5) IBM Research, MIT-IBM Watson AI Lab 6)Siemens AG, Munich, Germany

■ 背景

- 最近では、半帰納的、あるいは完全帰納的なシナリオを模索し、未知のエンティティや新たなエンティティに対する推論を可能にする研究が行われている
- しかし、これらのアプローチはすべてトリプルベースのKGのみを対象としており、超関係KG(例:Wikidata)についてはまだ適切なものがない

■ 手法

- 最近のグラフニューラルネットワークを利用した様々な半帰納的および完全帰納的リンク予測タスクを超関係性KGに適用することを調査する

■ 結果

- 超関係詞(hyper-relational facts)は帰納的な設定において、リンク予測(LP)の性能を大幅に向上させることを実証した

Model #QP	WD20K (33) SI				WD20K (25) SI					
	AMR(%)	MRR(%)	H@1(%)	H@5(%)	AMR(%)	MRR(%)	H@1(%)	H@5(%)	H@10(%)	
BLP 0	4.76	13.95	7.37	17.28	24.65	6.01	12.45	5.98	17.29	23.43
QBLP 0	7.04	28.35	14.44	28.58	36.32	6.75	17.02	8.82	22.10	29.50
QBLP 2	11.51	35.95	20.70	34.98	41.82	5.99	20.36	11.77	24.86	32.26
QBLP 4	11.38	<u>34.35</u>	<u>19.41</u>	<u>33.90</u>	<u>40.20</u>	12.18	21.05	12.32	24.07	30.09
QBLP 6	<u>4.98</u>	25.94	15.20	30.06	38.70	5.73	19.50	11.14	<u>24.73</u>	<u>31.60</u>

Model	#QP	WD20K (100) V1					WD20K (100) V2				
		AMR(%)	MRR(%)	H@1(%)	H@5(%)	H@10(%)	AMR(%)	MRR(%)	H@1(%)	H@5(%)	H@10(%)
BLP	0	22.78	5.73	1.92	8.22	12.33	36.71	3.99	1.47	4.87	9.22
CompGCN	0	37.02	10.42	<u>5.75</u>	15.07	18.36	74.00	2.55	0.74	3.39	5.31
QBLP	0	28.91	5.52	1.51	8.08	12.60	35.38	4.94	2.58	5.46	9.66
StarE	2	41.89	9.68	3.73	16.57	20.99	40.60	2.43	0.45	3.86	6.17
StarE	4	35.33	10.41	4.82	15.84	21.76	37.16	5.12	1.41	7.93	12.89
StarE	6	34.86	11.27	6.18	15.93	21.29	47.35	4.99	1.92	6.71	11.06
QBLP	2	18.91	10.45	3.73	16.02	<u>22.65</u>	28.03	6.69	3.49	<u>8.47</u>	12.04
QBLP	4	<u>20.19</u>	<u>10.70</u>	3.99	<u>16.12</u>	24.52	<u>31.30</u>	5.87	2.37	7.85	13.93
QBLP	6	23.65	7.87	2.75	10.44	17.86	34.35	<u>6.53</u>	<u>2.95</u>	9.29	13.13

半帰納的リンク予測

完全帰納的リンク予測

ISWC2021 サーベイ

富士通株式会社 小柳佑介

※資料内の図は原典からの引用です。

Session 2C: Semantic Integration

- **Best Paper Award** Leveraging Semantic Web technologies for digital interoperability in the European Railway domain (In-Use)
Julian Rojas, Marina Aguado, Polymnia Vasilopoulou, Ivo Velitchkov, Dylan Van Assche, Pieter Colpaert and Ruben Verborgh
- On constructing Enterprise Knowledge Graphs under quality and availability constraints (In-Use)
Matthew Kujawinski, Christophe Guéret, Chandan Kumar, Brennan Woods, Pavel Klinov and Evren Sirin
- Towards Semantic Interoperability in Historical Research: Documenting Research Data and Knowledge with Synthesis (in-Use)
Pavlos Fafalios, Konstantina Konsolaki, Lida Charami, Kostas Petrakis, Manos Paterakis, Dimitris Angelakis, Yannis Tzitzikas, Chrysoula Bekiari and Martin Doerr

Leveraging Semantic Web technologies for digital interoperability in the European Railway domain

Best Paper Award (In-Use)

Julián Andrés Rojas¹, Marina Aguado², Polymnia Vasilopoulou², Ivo Velitchkov³, Dylan Van Assche¹, Pieter Colpaert¹ and Ruben Verborgh¹

• ¹ IDLab, Department of Electronics and Information Systems, Ghent University – imec, Technologiepark-Zwijnaarde 122, 9052 Ghent, Belgium

• ² European Union Agency for Railways, Lille, France

• ³ DG DIGIT, Brussels, Belgium

European Union Agency for Railways(ERA) のデジタル相互運用性戦略の提案と、データ統合のためのセマンティックレイヤーを確立する基盤の構築

- 課題
 - 鉄道ドメイン内で相互作用する複数のアクター間のデータ交換は、従来のアプリケーション中心のアプローチに依存。デジタルの相互運用性への障壁が増え、メンテナンスと革新のコストが増加。
- 手法
 - (i) 28の参照データセットを含む、鉄道インフラストラクチャと認定車両型式のための公式オントロジー
 - (ii) 欧州の鉄道インフラを記述する再利用可能なKG
 - (iii) ユースケース開発の柔軟性が高く、コスト効率に優れたシステムアーキテクチャ
 - (iv) ルート互換性チェックをサポートするRDFネイティブなオープンソースWebアプリケーション。
- 結果／実績
 - ERAは公式に、データ交換のために、このアプローチとデータを採用。本番稼働に向け、ソリューションを開発中。

[video](#)

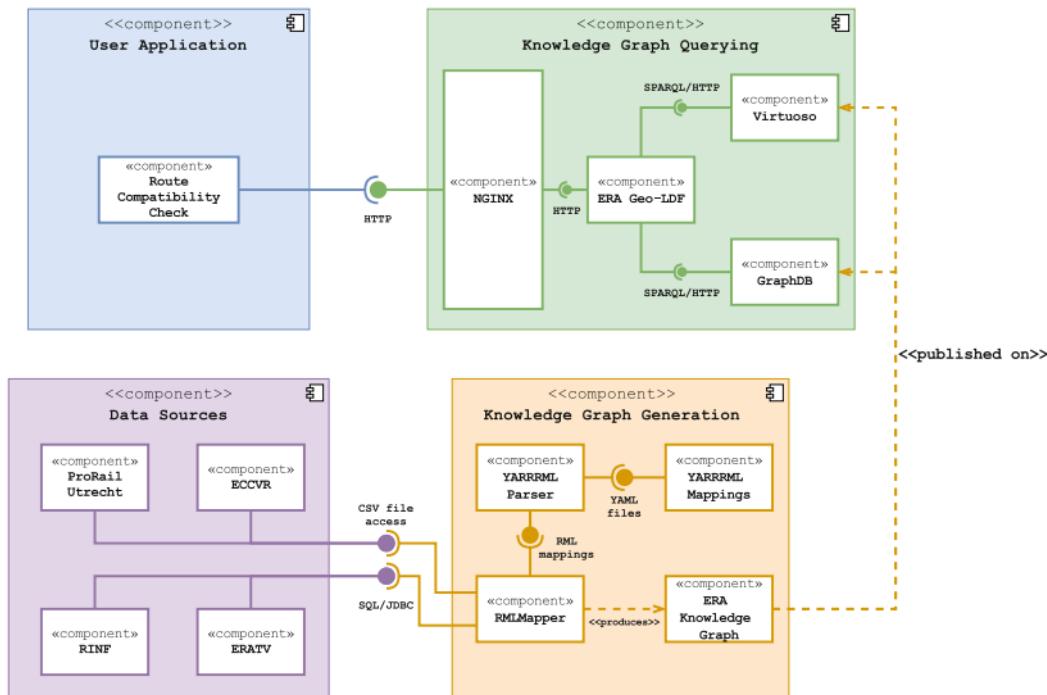


Fig. 3. Overview of the proposed solution architecture for semantic data interoperability across ERA's base registries.

On constructing Enterprise Knowledge Graphs under quality and availability constraints

(In-Use)

Matthew Kujawinski¹, Christophe Guéret¹, Chandan Kumar³, Brennan Woods³, Pavel Klinov² and Evren Sirin²

^{•1} Accenture Labs, San Jose, USA,

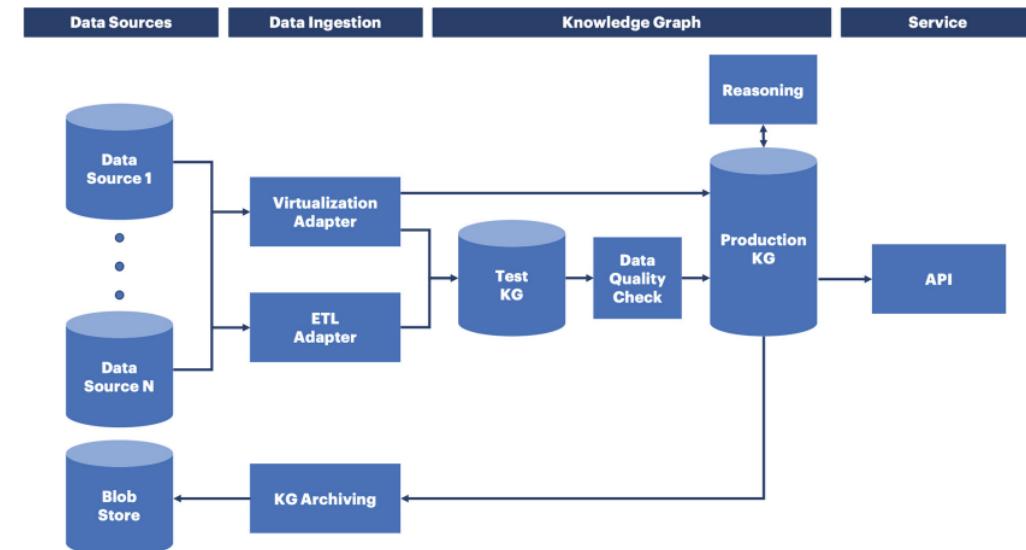
^{•2} Stardog, Arlington, USA

^{•3} Accenture, Dublin, Ireland

データ統合のリスク/課題に対処するためのアプローチの提案に向け、二つのクライアント事例を紹介

- 課題
 - データ統合にはセキュリティ、監査要件、品質管理など幾つかのリスクと課題が存在
- 手法
 - Virtualization adapter: Stardogでサポートされる言語SMSを使用し、データ源をKGへマッピング
 - ETL Adapter: ソースデータをTripleに変換するETLをApache NiFiで実現
 - Quality check: SHACLにより検証
 - Reasoning: Stardogビルトインのオントロジベースの推論機能
 - Data API: SPARQL/GraphQLを作成せずにデータにアクセスするAPI
 - KerberosとLDAPを使って、ユーザ認証を実現
- 結果／実績
 - 金融サービスのユースケース： PIIデータの流れの追跡
 - 製薬のユースケース： 分離したデータソースの統一ビューを製薬会社従業員に提供

[video](#)



Towards Semantic Interoperability in Historical Research: Documenting Research Data and Knowledge with *Synthesis* (In-Use)

Pavlos Fafalios¹, Konstantina Konsolaki¹, Lida Charalampidi¹, Kostas Petrakis¹, Manos Paterakis¹, Dimitris Angelakis¹, Yannis Tzitzikas^{1,2}, Chrysoula Bekiaris¹ and Martin Doerr¹

¹Centre for Cultural Informatics and Information Systems Laboratory, FORTH-ICS, Heraklion, Greece

²Computer Science Department, University of Crete, Heraklion, Greece

歴史学（美術史）の分野の研究プロジェクトにおいて使用する、研究データの統合管理システム
*Synthesis*と、歴史学者による活用事例

- 課題
 - 歴史学研究における、研究データ管理において、共同管理や、複数の多様な情報源のデータの統合、特定の文脈を越えたデータ再利用、などの課題
- 手法
 - *Synthesis*: Webベースの共同作業型のシステム
 - CIDOC-CRM、RDFなどの既存のstandardsを採用
 - XMLドキュメント内のデータをRDFデータに変換するプロセスを搭載。
 - X3MLフレームワークとX3MLマッピング定義言語を使用し、研究チームによる「データ入力」を、データエンジニアによる「オントロジーベースの統合」「知識ベース作成」から切り離す。
 - CIDOC-CRMへの変換により、より高度なクエリや分析を可能に。
- 結果／実績
 - 研究プロジェクトRICONTRANSにおいて活用。現在、5か国の10機関に属する40ユーザによって使用されている。エンティティの現在の(2021年7月2日現在)数は、1,089オブジェクト。
 - スキーマ設計や、語彙の制御などにおける教訓を得た。
 - プロジェクトは現在データ入力フェーズで、今後の普及と活用のために、地図可視化、データ公開、セマンティックネットワーク探索を計画中。

[video](#)

Session 4B: Information Extraction

- A Graph-based Approach for Inferring Semantic Descriptions of Wikipedia Tables (Research)
Binh Vu, Craig Knoblock, Pedro Szekely, Jay Pujara and Minh Pham
- Large-scale Multi-granular Concept Extraction Based on Machine Reading Comprehension (Research)
Siyu Yuan, Deqing Yang, Jiaqing Liang, Jilun Sun, Jingyue Huang, Kaiyan Cao, Yanghua Xiao and Rui Xie
- PCSG: Pattern-Coverage Snippet Generation for RDF Datasets (Research)
Xiaxia Wang, Gong Cheng, Tengteng Lin, Jing Xu, JeffZ. Pan, Evgeny Kharlamov and Yuzhong Qu

A Graph-based Approach for Inferring Semantic Descriptions of Wikipedia Tables

(Research)

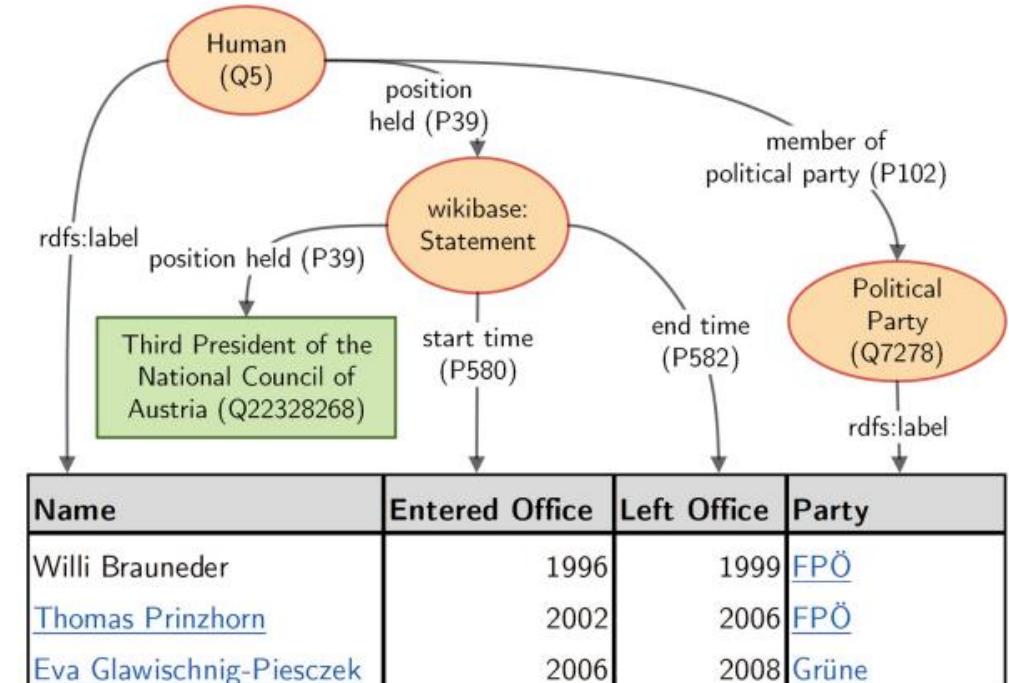
Binh Vu, Craig Knoblock, Pedro Szekely, Jay Pujara and Minh Pham

- USC Information Sciences Institute, Marina Del Rey, CA 90292, USA

Wikipedia中の表に対応するsemantic descriptionsを自動で構築する手法の提案

- 課題
 - semantic descriptionの構築は、多くの手作業を必要とする複雑なプロセスであり、手作業による誤りが発生しやすい
- 手法
 - 表内のハイパーリンクとWikidataを活用し、可能性のあるグラフを生成
 - そのグラフの関係に対し真・偽を推定
- 評価
 - SemTab 2020データセットでの実験で、既存手法よりも性能が向上
 - Wikipediaテーブルの大規模セットでの実験でもF1スコアにおいて28%も向上

[video](#), [github](#)



Large-scale Multi-granular Concept Extraction Based on Machine Reading Comprehension (Research)

Siyu Yuan¹, Deqing Yang¹, Jiaqing Liang², Jilun Sun², Jingyue Huang¹, Kaiyan Cao¹, Yanghua Xiao^{2,3} and Rui Xie⁴

^{1,3} School of Data Science, Fudan University, Shanghai, China

² School of Computer Science, Fudan University, Shanghai, China

³ Fudan-Aishu Cognitive Intelligence Joint Research Center, Shanghai, China

⁴ Meituan, Beijing, China

MRC-CE: エンティティのテキストから概念を抽出するための、概念抽出フレームワーク

- 課題

- 既存のKGは、細かい概念をカバーできていない

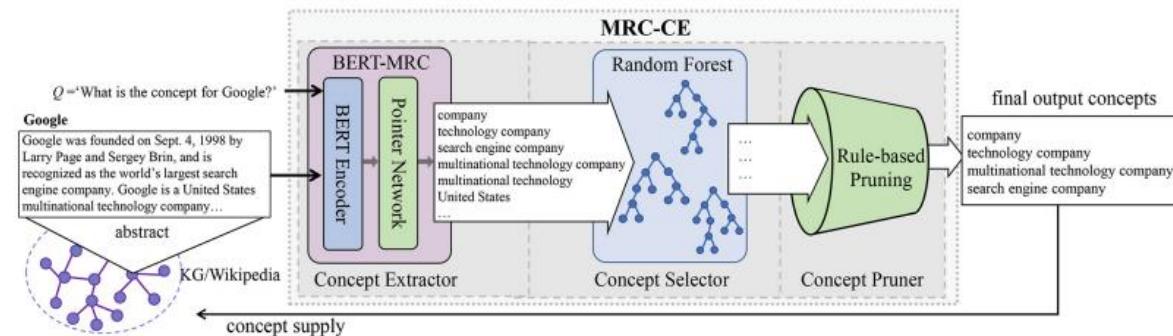
- 手法

- BERT Encoder+Pointer Networkにより、細かい粒度の概念を抽出するモデルを構築
- ランダムフォレストとルールベースの枝刈りにより、精度と再現率を向上

- 評価

- 多言語KG（英語Probbaseと中国語CN-DBpedia）での評価実験において、KGCにおけるMRC-CEの優位性を確認
- 特に、CN-DBpediaにおいて、エンティティごとにMRC-CEを実行することで、7,053,900以上の新しい概念(instanceOf関係)を抽出

[github](#)



PCSG: Pattern-Coverage Snippet Generation for RDF Datasets

Xiaxia Wang¹, Gong Cheng¹, Tengteng Lin¹, Jing Xu¹, JeffZ. Pan², Evgeny Kharlamov^{3,4} and Yuzhong Qu¹ (Research)

¹ State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

³ Bosch Center for Artificial Intelligence, Robert Bosch GmbH, Gerlingen, Germany

² School of Informatics, University of Edinburgh, Edinburgh, UK

⁴ Department of Informatics, University of Oslo, Oslo, Norway

RDFデータセットのスニペット生成手法の提案

- 課題

- RDFデータセットを活用するにあたって構造の理解は重要だが、理解は困難

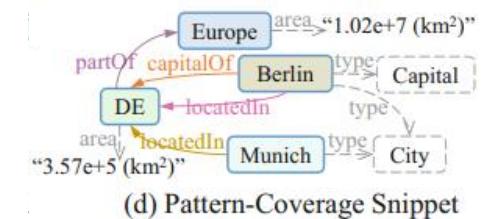
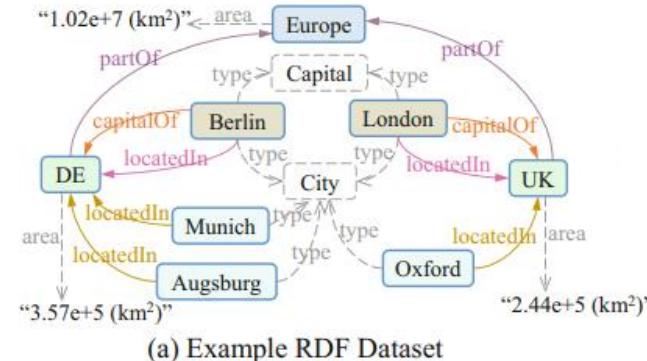
- 手法

- 基本： group steiner tree problemを解くことで、全てのパターンをカバーするコンパクトなスニペットを生成。
- PCSG： 各要素から抽出した最小数のサブスニペットをマージして生成。集合被覆問題を解くことで実現。
- QPCSG： クエリバイアスをかけたスニペット生成。クエリキーワードとマッチングするエンティティをカバーするスニペットを生成。

- 評価

- クエリ・スキーマのカバレッジ、ユーザ好みにおいて、既存の手法より良い結果
- スペースの節約・実行時間の実用性を実証

[video](#), [github](#)



[Session 3A – Knowledge Graph Embeddings]

Improving Knowledge Graph Embeddings with Ontological Reasoning (Research)

- 既存のKG embedding手法を使い、オントロジーと矛盾する負例を動的に与えることで、精度向上させた

背景 & 動機

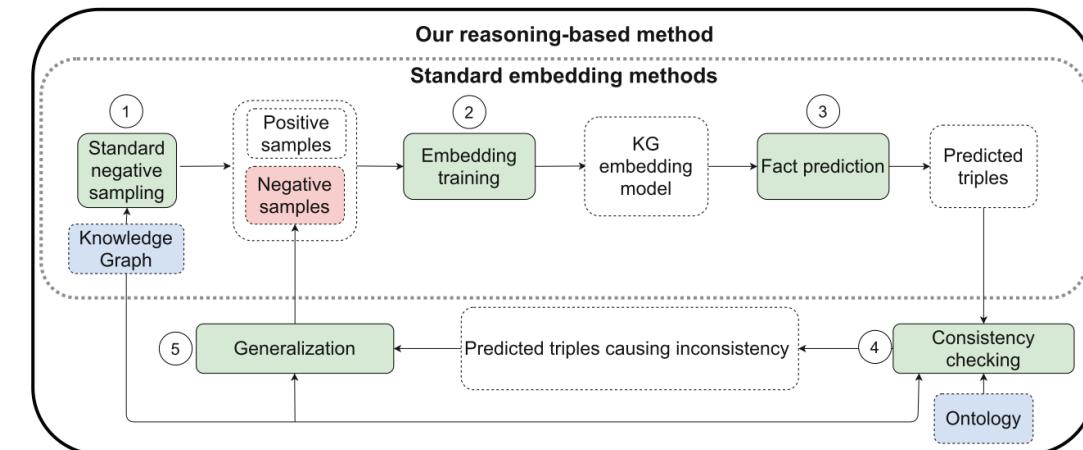
- KG completionのタスクにembeddingの手法を適用するにあたり、負例をどう与えるかが問題とされてきた
- Local closed world assumptionに基づく典型的な方法では、生成した負例が実は正例の場合もある
- オントロジーを使って静的に事前に負例を生成する先行研究があるが、得られたモデルで誤って正しいと推定した負例を含んでいない場合あり
- そこで、オントロジーを活用して、本当の負例のみを動的に生成したい

手法 (右上図の通り; 4と5がポイント)

- 4で、推定されたトリプルがオントロジーと矛盾していたらそれを負例とする
- 5で、主語や述語を、同種 (詳細略) のエンティティに置換えたトリプルも負例とする

結果

- 既存の手法としてTransEとComplExを使い、精度向上を確認
- 静的な先行研究と比較して概ね精度向上を確認
- オントロジーと矛盾するトリプルが上位の推定結果に出るかという観点での実験では概ね良い結果 (右下図)



Model	KG	Prediction	Default Training		Static Sampling		ReasonKGE	
			Inc@1	Inc@10	Inc@1	Inc@10	Inc@1	Inc@10
TransE	LUBM3U	subject	0.169	0.270	0.428	0.250	0.037	0.133
		object	0.095	0.097	0.212	0.104	0.005	0.007
	YAGO3-10	subject	0.075	0.280	0.629	0.492	0.075	0.273
		object	0.026	0.136	0.114	0.089	0.020	0.117
ComplEx	LUBM3U	subject	0.311	0.652	0.401	0.663	0.217	0.585
		object	0.413	0.538	0.428	0.544	0.170	0.460
	YAGO3-10	subject	0.041	0.097	0.177	0.136	0.036	0.069
		object	0.008	0.012	0.003	0.007	0.005	0.007
DBpedia15K	LUBM3U	subject	0.113	0.198	0.169	0.128	0.071	0.143
		object	0.037	0.115	0.065	0.084	0.015	0.074
	YAGO3-10	subject	0.488	0.667	0.436	0.695	0.344	0.583
		object	0.397	0.585	0.365	0.528	0.318	0.533

[Session 3A – Knowledge Graph Embeddings] Using Compositional Embeddings for Fact Checking (Research)

- KGにおけるFact Checkingのタスクに、初めてembeddingsを適用し、精度向上させた

• 背景

- 同じ研究室の先行研究 [Syed2019] (ISWC2019で発表; 右図) をベースとする

• 動機

- 膨大なパスの探索が問題→探索するパスのembeddingとチェックしたいfactの述語のembeddingの近さを使ってうまくパスの探索をすることで、効果的により長いパスも探索できるようにならないか?

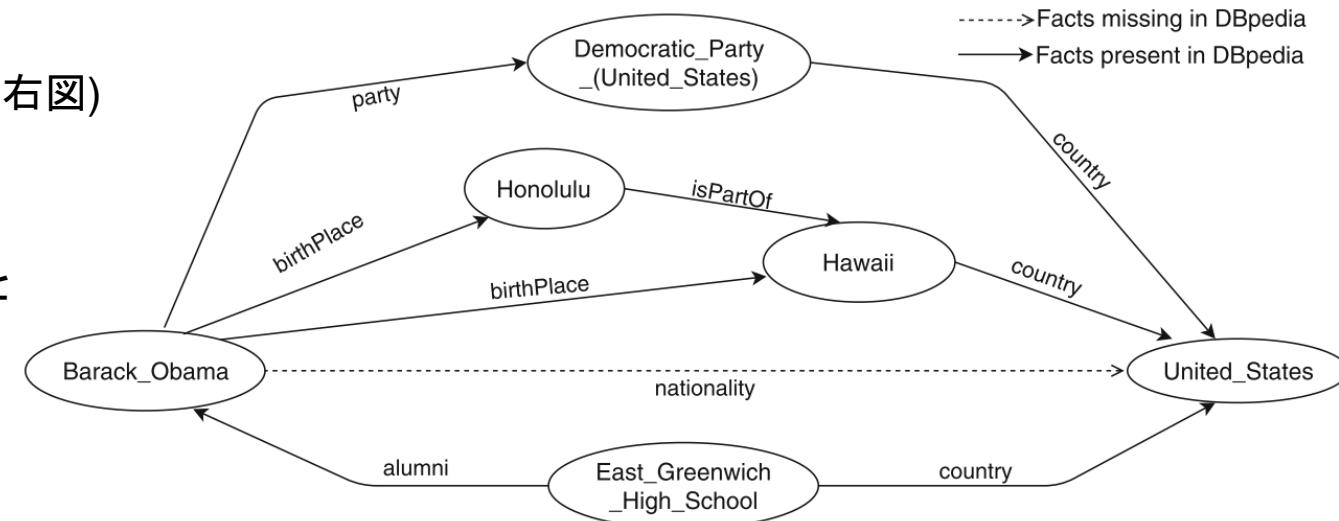
• 手法

- (1) パスの探索 (embeddingsを利用): 最も簡単なKG embeddingモデルTransEの場合、ベクトルを足しつつA*アルゴリズムで上位N個のパスを求める
- (2) パスのスコア付け: 先行研究のマイナーチェンジでPNPMIを用いて算出する
- (3) ファクトのスコア付け: 先行研究と異なり、全てのパスのPNPMI値を3乗平均する

• 結果

- 既存の10手法と比較実験し、概ね良い結果 (一部負けている)
- 既存手法と本手法を組合せた実験をしたところ、全10手法において精度向上

- **報告者感想:** (2) のステップで何でembeddingsを使わないのか (試したけどうまくいかなかった??)



[Session 3A – Knowledge Graph Embeddings]

LiterallyWikidata - A Benchmark for Knowledge Graph Completion using Literals (Resources)

- リテラル (数値、時刻、文字列) も含むKG completion用の新しいベンチマークデータセットを作った
- リソースの設計方針
 - WikidataとWikipediaを入力とし、サイズの異なる3つのデータセットを作成: LitWD1K, LitWD19K, LitWD48K
 - 手順: (1) リテラル数の多いエンティティをシードとし1か2ホップ辺り (2) 次数が高いエンティティを残し (3) 良くないデータを削る
- 新規性
 - 既存の数多くのデータセット (このうち比較実験対象としたのはFB15K-237, CoDEx-M) と比べて、次の条件を全て満たす
 - リテラル (数値、時刻、文字列) を含む
 - Skewness: 主語や目的語に偏りのある述語が少ない
 - Symmetricity, Inversion: 対称的な述語や逆の関係にある述語がない
 - Cartesian Product or Fixed-set Relations: ある集合の主語とある集合の目的語間に必ず同じ述語関係があるということがない
 - 手動作業が少ない (新しいデータセットも容易に作成可能)
- 品質
 - KG embeddingのモデルDistMultとDistMultLiteralの実験結果を比較すると後者の方が良い→リテラルが有効活用
 - サイズ的に同等のFB15K-237, CoDEx-MとLitWD19Kの実験結果を比較すると後者の方が悪い→難易度の高いデータセット
- リソースの可用性: [データセット](#)、[プログラム](#)
- 次にすべきこと
 - (特に) クラウドソーシング的に作られたWikiDataを元としたことによる、データのバイアスの調査

ご協力いただき
ありがとうございました！