

人工知能学会SWO研究会ワークショップ

2020/9/7(水)

ナレッジグラフ推論チャレンジ2020技術勉強会

ナレッジグラフ 推論チャレンジ

Knowledge Graph Reasoning Challenge

ナレッジグラフ推論チャレンジの紹介
～開催概要・提供するナレッジグラフ～



講演内容

▶ ナレッジグラフ推論チャレンジの紹介

➤ 開催趣旨

➤ 第3回ナレッジグラフ推論チャレンジ2020の概要

▶ 提供するナレッジグラフ

➤ スキーマの概要

➤ 基本的な情報の検索方法

▶ 対象とする小説のあらすじ



推論チャレンジとは？

◆ シャーロック・ホームズのような

“推理”（推論）ができるAIシステムの開発
を目指した技術コンテスト

➡ チャレンジのねらい

➤ **説明可能性（解釈可能性）**を有するAI技術に関する最新技術の促進・共有と、その分析・評価、体系化を行う。

➡ チャレンジタスク

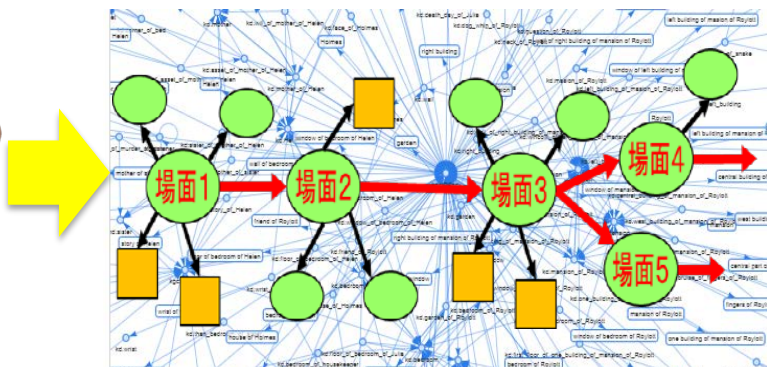
➤ **推理小説のナレッジグラフ**（ホームズの短編小説）を対象に、ホームズと同じ結論に辿り着き、その理由を説明する。



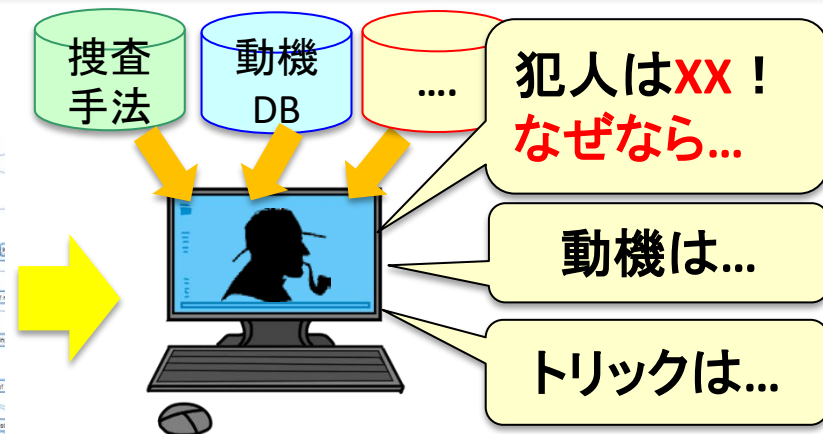
チャレンジの概要



ホームズ
の推理小説



ナレッジグラフ（知識グラフ）
としてデータ化



さまざまな知識/手法を用いて
事件の真相を推理し、理由を
説明するAIシステムの開発

推理(推論)の例

➤ ジュリアの死に不審な点があると、ヘレンから相談を受ける

↓ ←犯人は現場にいた人【外部知識】

➤ 「ジュリアが死んだ日」に現場の屋敷にいた人物の一覧を
取得する



チャレンジの狙い

- ▶ **説明可能性(解釈可能性)**を有するAI技術に関する最新技術の促進・共有と, その分析・評価, 体系化を行う.
- ▶ 特に, 現実社会を反映したより複雑な, 例えば時間的, 因果関係的, 確率的関係性を含む問題を扱うため, **帰納的な機械学習(推定)**と**演繹的な知識活用(推論)**を融合したAI技術を対象とする.



チャレンジタスク

- ➡ **推理小説のナレッジグラフ**（シャーロックホームズの短編小説）を対象に，ホームズと同じ結論に辿り着き，その理由を説明する。
- ➡ **なぜ，推理小説を対象にするのか？**
 - 現実社会の複雑な関係性を含みながら**仮想的にクローズ**な（答えがあり，それに至る制約を制御できる）タスクを設計できる
 - 読者が納得しないと小説として成立しないという**人間に対する説明性**を有している
 - タスクによっては，不確実情報や証拠写真など確率的な処理や機械学習を入れないと解けなかったり，陽に書かれていない**常識知識を補完**しなくては解けない等，**推定と推論の融合**を促せる
 - 小説が**広く一般に知られており関心を引きやすい**，など
- ➡ **第4回チャレンジ以降は社会問題解決に関するベストプラクティス集なども検討中**



- ➡ コロナの影響で国際版チャレンジ併催を企画していたJISTが延期...
- ➡ 6/9より、**第3回国内版推論チャレンジ2020**の作品応募を開始！
 - 既存の**5つのナレッジグラフ**の洗練化版
まだらのひも、踊る人形、背中の曲がった男、
悪魔の足、花婿失踪事件
 - 新規に**3つのナレッジグラフ**を追加(9月公開予定)
僧坊荘園、入院患者、白銀号殺人事件
 - スケジュール
 - ✓ 2020/06/09 応募開始
 - ✓ 2020/09/xx 追加KG公開予定
 - ✓ 2020/10/31 **チャレンジ応募×切**
 - ✓ 2020/12/末 最終審査会・授賞式(状況を見てオンライン開催を検討)
- ➡ 新審査基準：
可能な限り複数の作品に共通的に適用できる手法を高く評価します！



➡ 対象とするKG

- 8つの小説のうち、いずれの小説を対象にしてもよい
(どれか1つだけ, 2つだけ…などでもOK)
- できるだけ多くの小説が、同じシステム(仕組み)で解けるとよい
- 各小説で使用するKGの範囲を段階的に変える(昨年同様)
→ 完全(すべてのKG) / 不完全(10%) / 不完全(25%)
- ナレッジグラフの独自拡張も可能(昨年同様)

➡ 対象とするタスク

- ① **本部門**: 対象小説1つ以上のタスクを解くシステムを開発
- ② **ツール部門**: いずれかのタスクを部分的に解くツールを開発
例) 容疑者の推定, アリバイ検証, 動機説明, など
★「自然言語文をトリプル化」するKG構築支援ツールの応募も可
- **アイデア部門**: ①, ②の実現方法のアイデア (実装なしでOK)



講演内容

▶ ナレッジグラフ推論チャレンジの紹介

➤ 開催趣旨

➤ 第3回ナレッジグラフ推論チャレンジ2020の概要

▶ 提供するナレッジグラフ

➤ スキーマの概要

➤ 基本的な情報の検索方法

▶ 対象とする小説のあらすじ

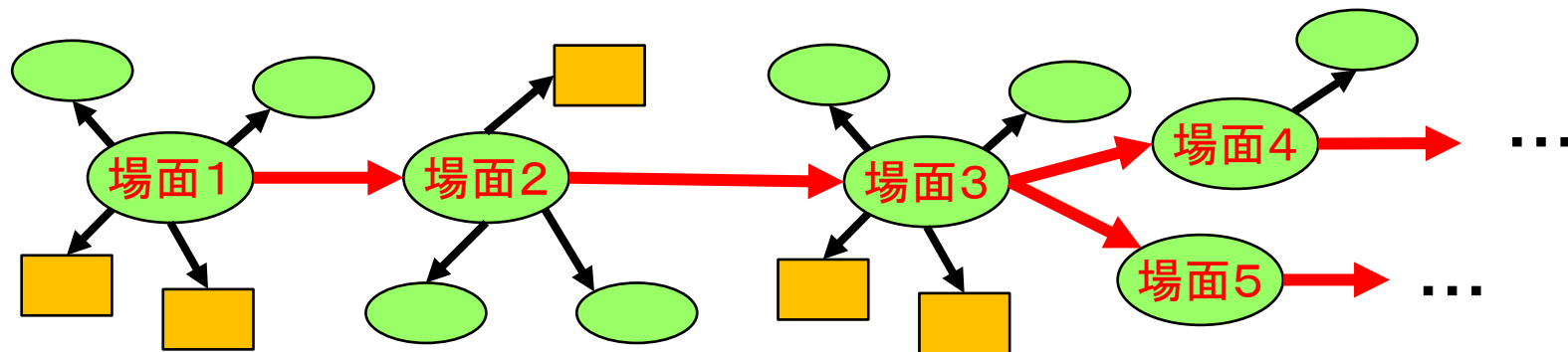


➡ ナレッジグラフの要求仕様

- 犯人を推論(推理)するのに必要な知識を提供する
- 「推理小説」で描かれる様々な状況を, できるだけ統一的な形式で計算機処理(検索・推論・etc.)可能にする

➡ ナレッジグラフ化の基本方針

- 「推理小説」の内容を, 最小単位の「**場面(シーン)**」に分割
→ 場面ごとにID(IRI)を付与
- 「**各場面の記述内容**」および「**場面間の関係**」をグラフ化
→ グラフ化に必要なクラス・プロパティを定義

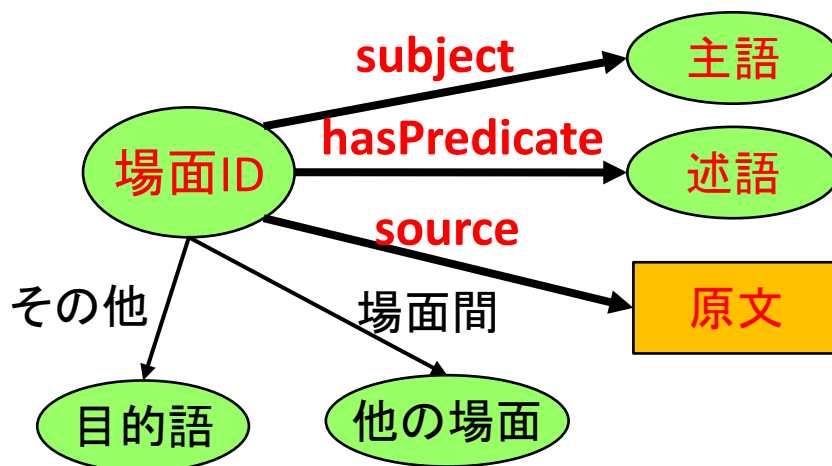




場面(シーン)のスキーマ

➤ 場面を表現するプロパティ

- **subject**: その場面の記述において主語となる人や物
- **hasPredicate**: その場面の内容を表す述語
- **場面の詳細を表す目的語**: whom(だれに), where(どこで), when(いつ), what(何を), how(どのように), ...etc.
- **場面間の関係**: then, if, because, ...etc.
- **time**: その場面が起こった絶対時間(xsd:DateTime)
- **source**: その場面の原文(英語/日本語のリテラル)



主語 → 述語 → 目的語

の形でないことに注意
→ 場面に関わる複数の
情報をまとめるため
場面ID中心に記述

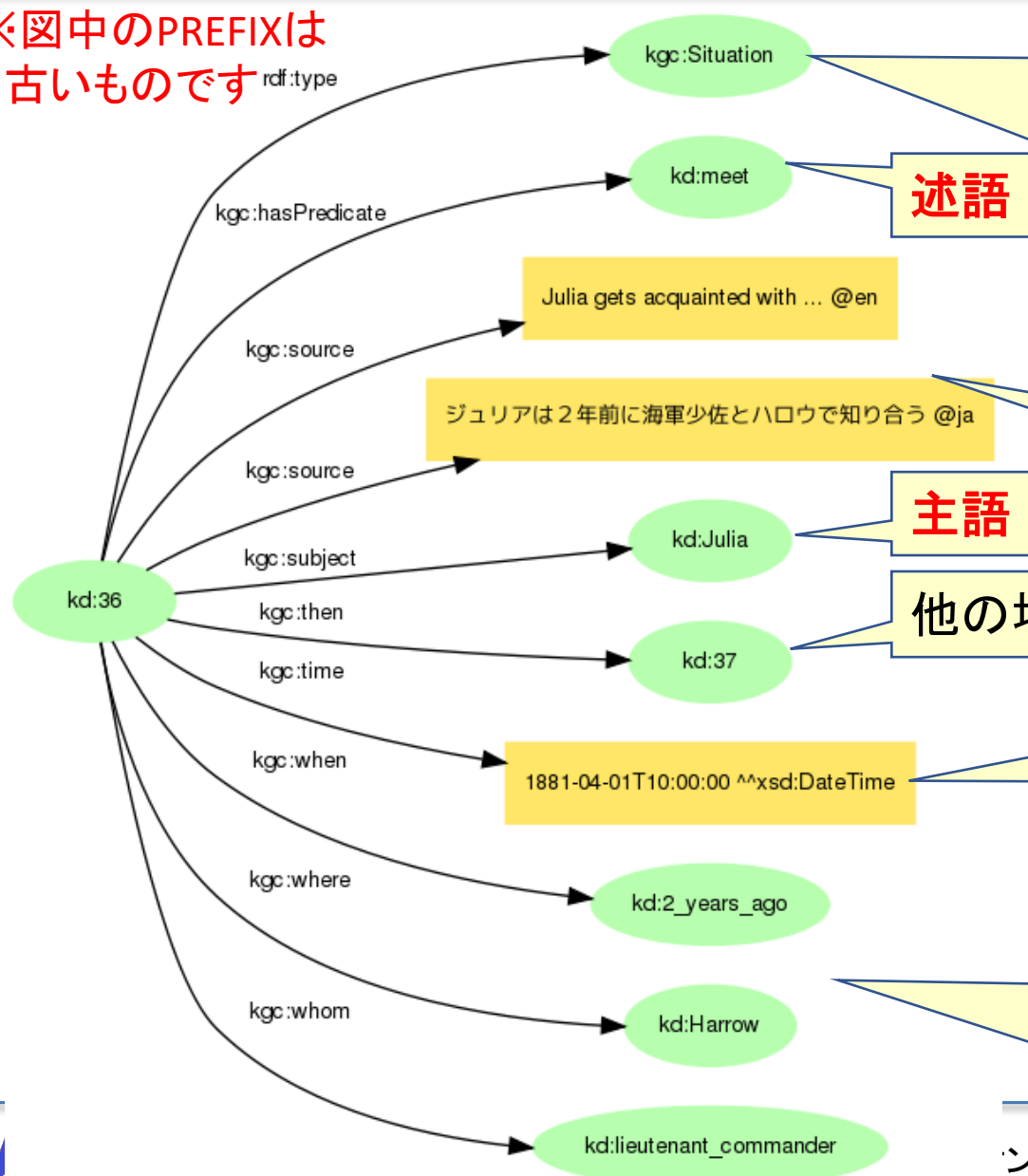


- ➡ 以下，提供するナレッジグラフの記述例，および，それを踏まえたSPARQLクエリによる検索例を示す。
- ➡ クエリのサンプルは，下記に公開している
 - <https://github.com/KnowledgeGraphJapan/LOD-ws-2020/blob/master/kgrc2020ws/SPARQL-Sample-KGRC2020.md>
 - これをコピーして
<http://knowledge-graph.jp/sparql.html>
のクエリ欄に入力することで実行できる



場面(シーン)スキーマ 記述例

※図中のPREFIXは
古いものです rdf:type



述語

場面の種類（クラス）分け

Scene : 上位クラス

-**Situation** : 事実・状況の描写

-**Statement** : Aの発言

-**Talk** : AのBへの発言

-**Thought** : Aの考え

原文（英語/日本語）

主語

他の場面

絶対時間※小説内に基準日時を設定している

主語・述語・目的語は全て「リソース」として定義
→他の場面で同じ目的語を参照可能



▶ 「まらだのひも」の場面36の**全トリプル(述語・目的語)**を取得する

```
1 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
2
3 SELECT DISTINCT *
4 FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
5 WHERE {
6     <http://kgc.knowledge-graph.jp/data/SpeckledBand/36> ?p ?o .
7 }
```



▶ 「まらだのひも」の場面36のhasPredicate(その場面の
内容を表す述語)を取得する

```
1 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
2
3 SELECT DISTINCT *
4 FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
5 WHERE {
6     <http://kgc.knowledge-graph.jp/data/SpeckledBand/36> kgc:hasPredicate ?o .
7 }
```



▶ 「まらだのひも」の場面36のsubject(その場面の主語)を取得する

```
1 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
2
3 SELECT DISTINCT *
4 FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
5 WHERE {
6     <http://kgc.knowledge-graph.jp/data/SpeckledBand/36> kgc:subject ?o .
7 }
```



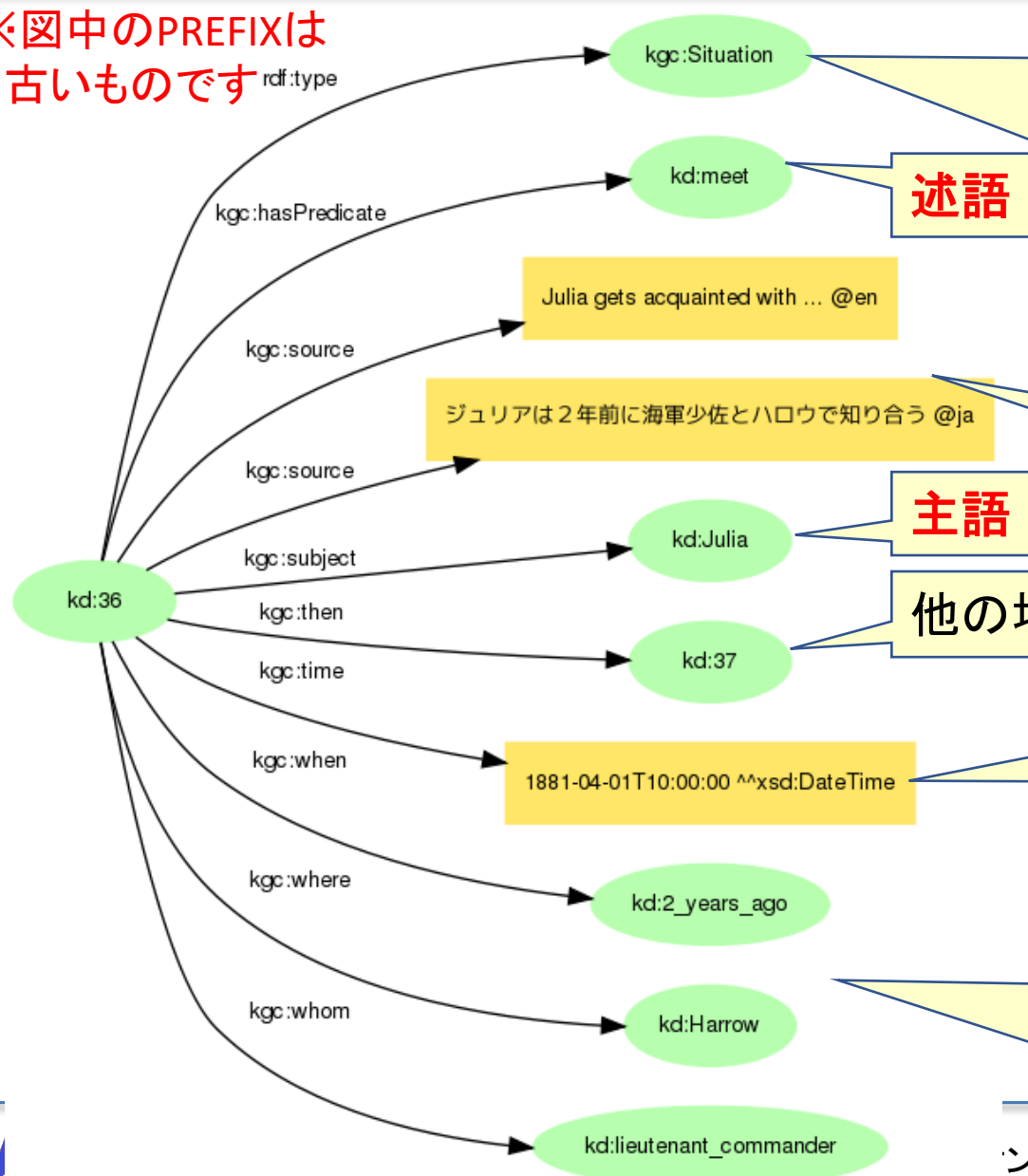
- ▶ 「まらだのひも」で用いられているhasPredicate(その場面の内容を表す述語)の一覧を取得する

```
1 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
2
3 SELECT DISTINCT ?o
4 FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
5 WHERE {
6     ?s kgc:hasPredicate ?o .
7 }
```



場面(シーン)スキーマ 記述例

※図中のPREFIXは
古いものです rdf:type



場面の種類 (クラス) 分け

Scene : 上位クラス

-**Situation** : 事実・状況の描写

-**Statement** : Aの発言

-**Talk** : AのBへの発言

-**Thought** : Aの考え

述語

原文 (英語/日本語)

主語

他の場面

絶対時間※小説内に基準日時を設定している

主語・述語・目的語は全て「リソース」として定義
→他の場面で同じ目的語を参照可能



➡ 「まらだのひも」の **Situation** (事実・状況の描写) の **一覧** を取得する

```
1 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
2
3 SELECT DISTINCT ?o
4 FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
5 WHERE {
6     ?s kgc:hasPredicate ?o .
7 }
```



➡ 「まらだのひも」の**全場面の一覧**を取得する

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
4
5 SELECT DISTINCT *
6 FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
7 WHERE {
8     ?s rdf:type/rdfs:subClassOf kgc:Scene .
9     ?s rdf:type ?o .
10 }
```



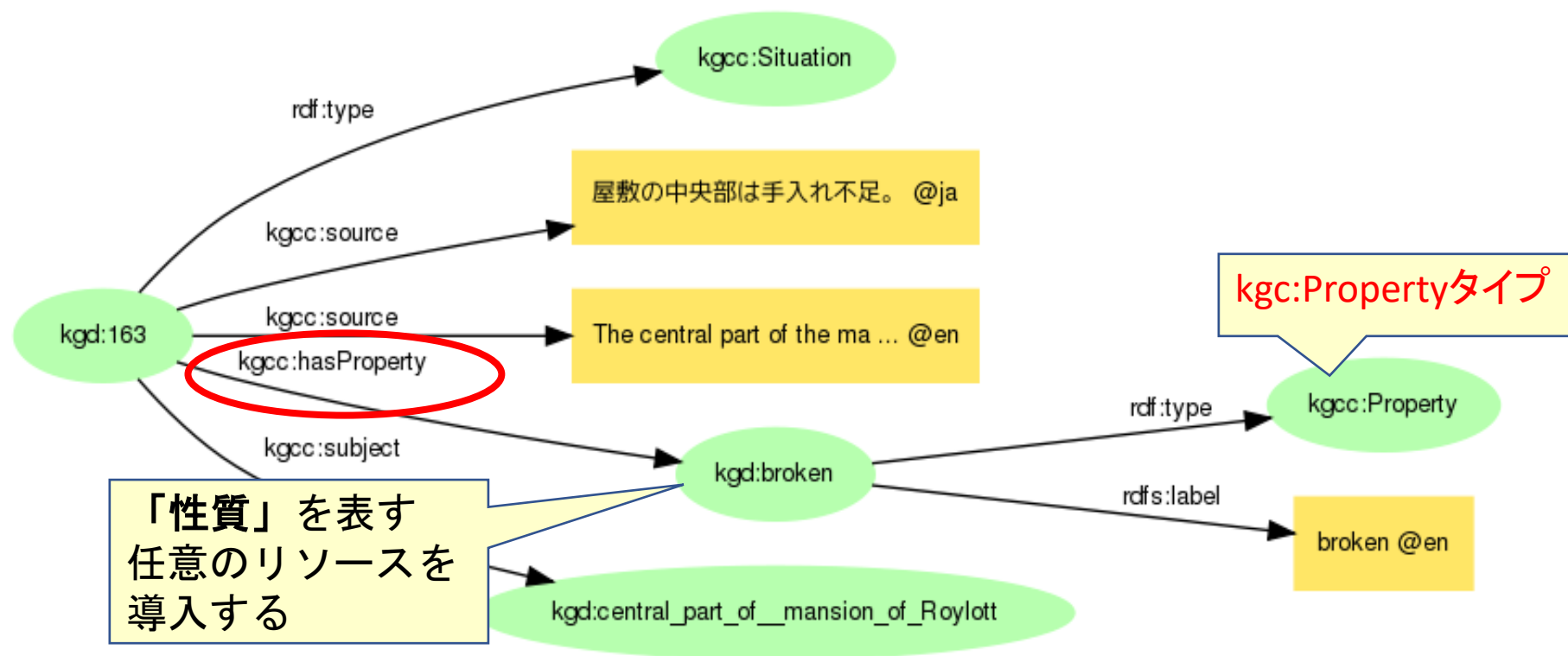
➡ 「全小説」で述語がmeetの場面の一覧を取得する

```
1  PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
2
3  SELECT DISTINCT *
4  FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
5  FROM <http://kgc.knowledge-graph.jp/data/DancingMen>
6  FROM <http://kgc.knowledge-graph.jp/data/ACaseOfIdentity>
7  FROM <http://kgc.knowledge-graph.jp/data/DevilsFoot>
8  FROM <http://kgc.knowledge-graph.jp/data/CrookedMan>
9
10 WHERE {
11     ?s kgc:hasPredicate <http://kgc.knowledge-graph.jp/data/predicate/meet> .
12 }
```



性質・状態(hasProperty)の例

※図中のPREFIXは
古いものです





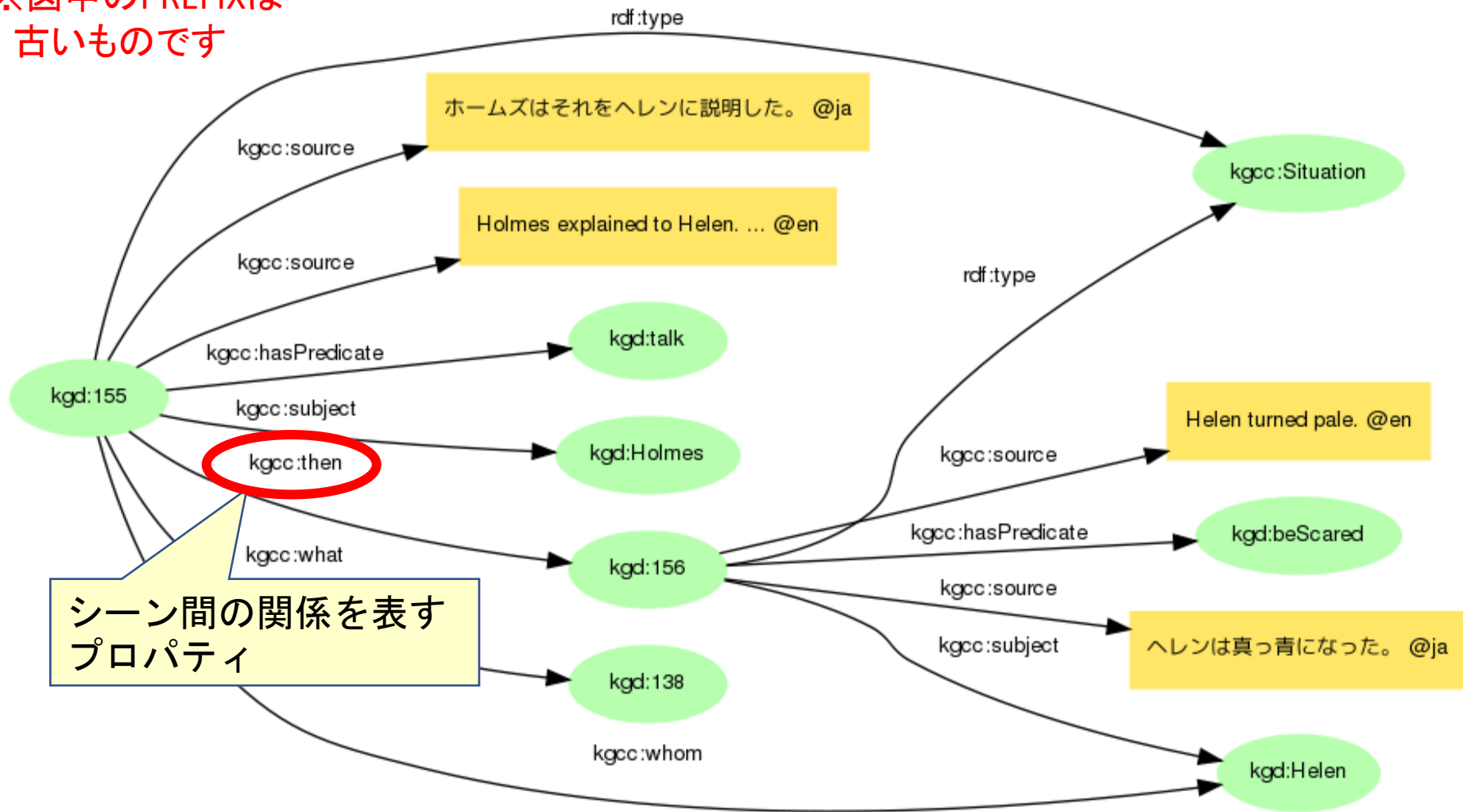
- ▶ 「まらだのひも」で用いられているhasProperty(その場面の内容を表す状態)の一覧を取得する

```
1 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
2
3 SELECT DISTINCT ?o
4 FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
5 WHERE {
6     ?s kgc:hasPredicate ?o .
7 }
```




シーン間の関係

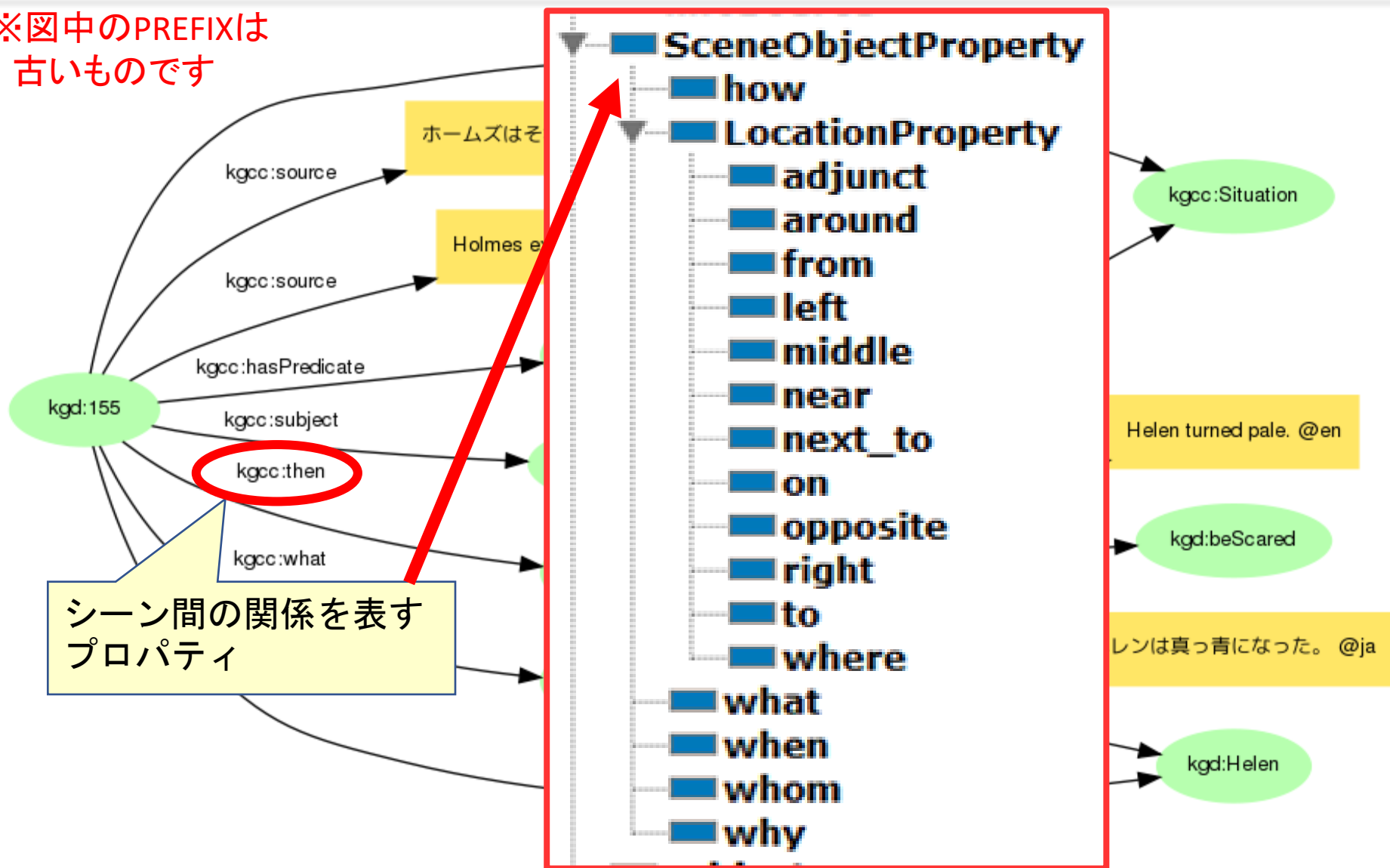
※図中のPREFIXは
古いものです





シーン間の関係の種類

※図中のPREFIXは古いものです





➡ 「まらだのひも」で**場面間のつながりの一覧**を取得する

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
4
5 SELECT DISTINCT *
6 FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
7 WHERE {
8     ?s rdf:type/rdfs:subClassOf kgc:Scene .
9     ?o rdf:type/rdfs:subClassOf kgc:Scene .
10    ?s ?p ?o .
11 }
```



各語彙のIRIの決め方

- ➡ 対象とする小説が増えたことで、異なる小説を横断した処理ができるようにIRIの決め方を統一
- ➡ 各小説ごとに「ベースIRI」および「グラフIRI」
 - <http://kgc.knowledge-graph.jp/data/**小説名**/>
 - 例:「踊る人形」のシーン100は
<http://kgc.knowledge-graph.jp/data/**DancingMen**/100>
- ➡ 小説をまたいで共通化して参照する語彙のIRI
 - hasPredicateで参照する「述語」、および、hasPropertyで参照する「性質・状態」
→<http://kgc.knowledge-graph.jp/data/**predicate**/XXX>
 - 固有名詞(例:ホームズ, 地名)
→<http://kgc.knowledge-graph.jp/data/**YYY**>



ナレッジグラフの公開

➡ ナレッジグラフの公開場所

- 「推論チャレンジ」のサイト→ナレッジグラフの公開

<https://github.com/KnowledgeGraphJapan/KGRC-RDF>

- 今年(2020年)版は、**2020**のフォルダ

<https://github.com/KnowledgeGraphJapan/KGRC-RDF/tree/master/2020>

➡ 公開形式

- 小説ごとのRDFファイル(Turtle形式)のダウンロード

- SPARQLエンドポイント

<http://kg.hozo.jp/fuseki/kgrc/sparql> (APIとして使う)

<http://knowledge-graph.jp/sparql.html> (ブラウザから)

- 可視化ツール

<http://knowledge-graph.jp/visualization/>



推理小説のナレッジグラフ

SWO研究会・勉強会での予備的作業を経て、有志数名でナレッジグラフ化

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
PREFIX kd: <http://kgc.knowledge-graph.jp/data/SpeckledBand/>
SELECT ?s ?p ?o
WHERE {
  ?s ?p ?o .
  filter(?p != rdfs:comment)
  filter(?p != kgc:source)
  filter(?o != kgc:Situation)
  filter(?o != kgc:Action)
  filter(?o != kgc:Object)
}
limit 1000
    
```

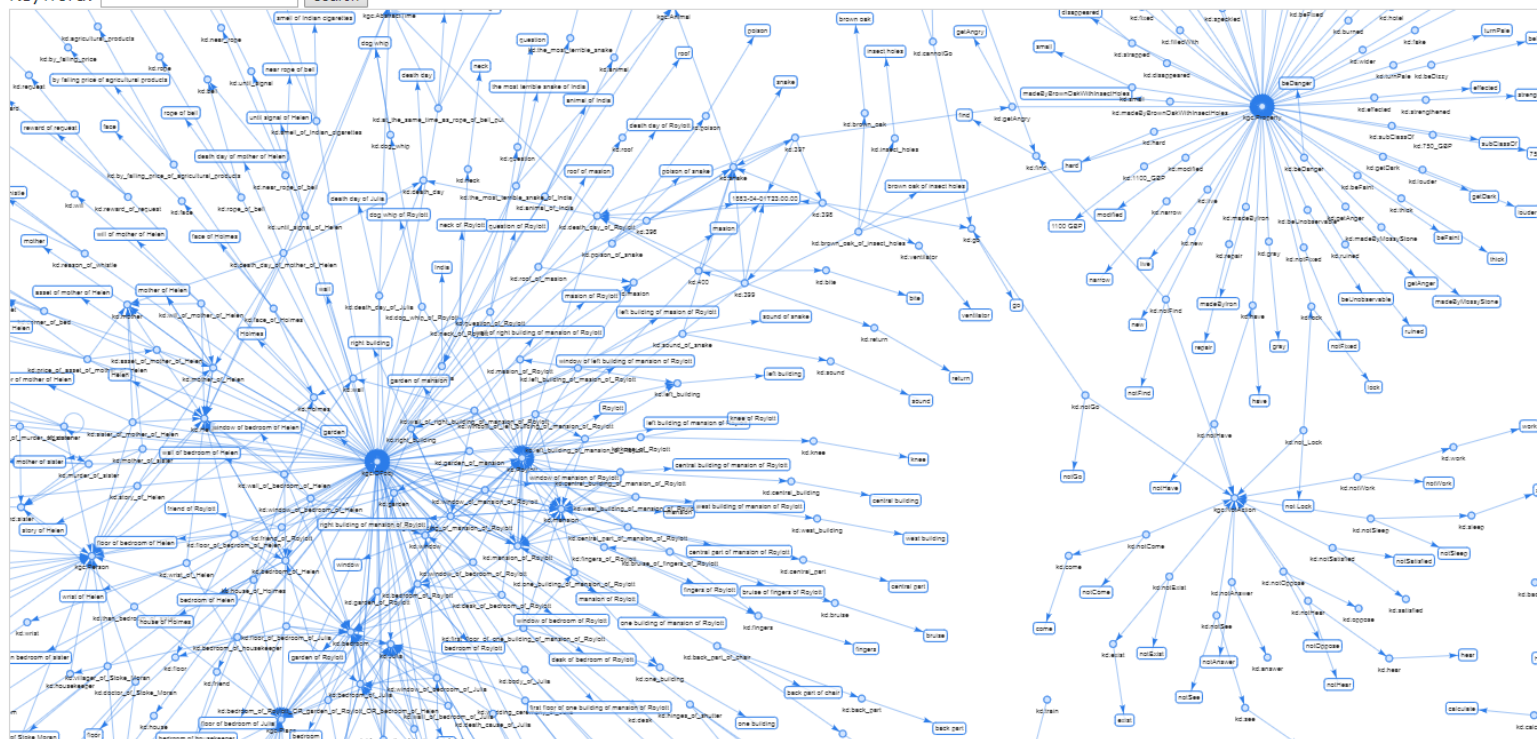
クエリ言語SPARQLによる検索

SPARQL Endpoint: <http://lod.hozo.jp/repositories>

グラフDB(キーワード検索也可)

ナレッジグラフ(RDF形式)

Keyword:



<http://knowledge-graph.jp/visualization/>



講演内容

▶ ナレッジグラフ推論チャレンジの紹介

➤ 開催趣旨

➤ 第3回ナレッジグラフ推論チャレンジ2020の概要

▶ 提供するナレッジグラフ

➤ スキーマの概要

➤ 基本的な情報の検索方法

▶ 対象とする小説のあらすじ