

場面展開を中心とした推理小説ナ レッジグラフの可視化手法

大阪電気通信大学
情報通信工学部情報工学科4回生
片桐 朋哉

研究概要

背景

- AIにおける説明可能性を高めるコンテストとして、人工知能学会セマンティックウェブとオントロジー研究会が開催する「ナレッジグラフ推論チャレンジ」がある。
- 「ナレッジグラフ推論チャレンジ」では推理小説をグラフ構造にしたナレッジグラフ(以下KG)を利用して機械的に推論を行う。

KG(ナレッジグラフ)

- ナレッジグラフとは、データ同士の知識の繋がりをグラフ構造によって表現人間にも機械にも可読しやすいようデータ形式が構造化
- コンテストより提供の小説ナレッジグラフにおいてはRDF(Resource Description Framework)に従って構造化

RDF

RDFによる構造化とは

- RDFによって構造化されたデータは、主語、述語、目的語のトリプルと呼ばれる3つ組みによって、データとデータ、その繋がりが表現されている。

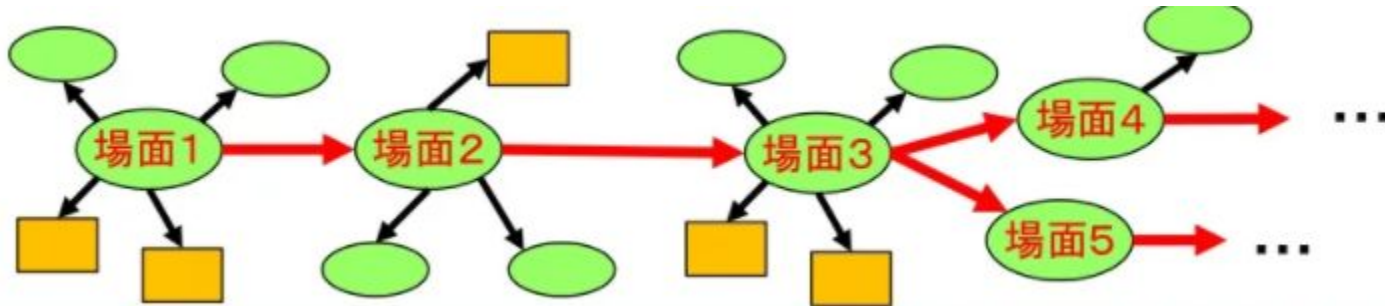


図1 RDFの構造化されたデータ

研究概要

目的

- しかし現状「ナレッジグラフ推論チャレンジ」で公式によって提供されている閲覧ツールと可視化ツールがあるのだが、場面間の関係やつながりが見にくい状態となっている。そこで、一目で文章の流れが分かるような可視化ツールを作成する。

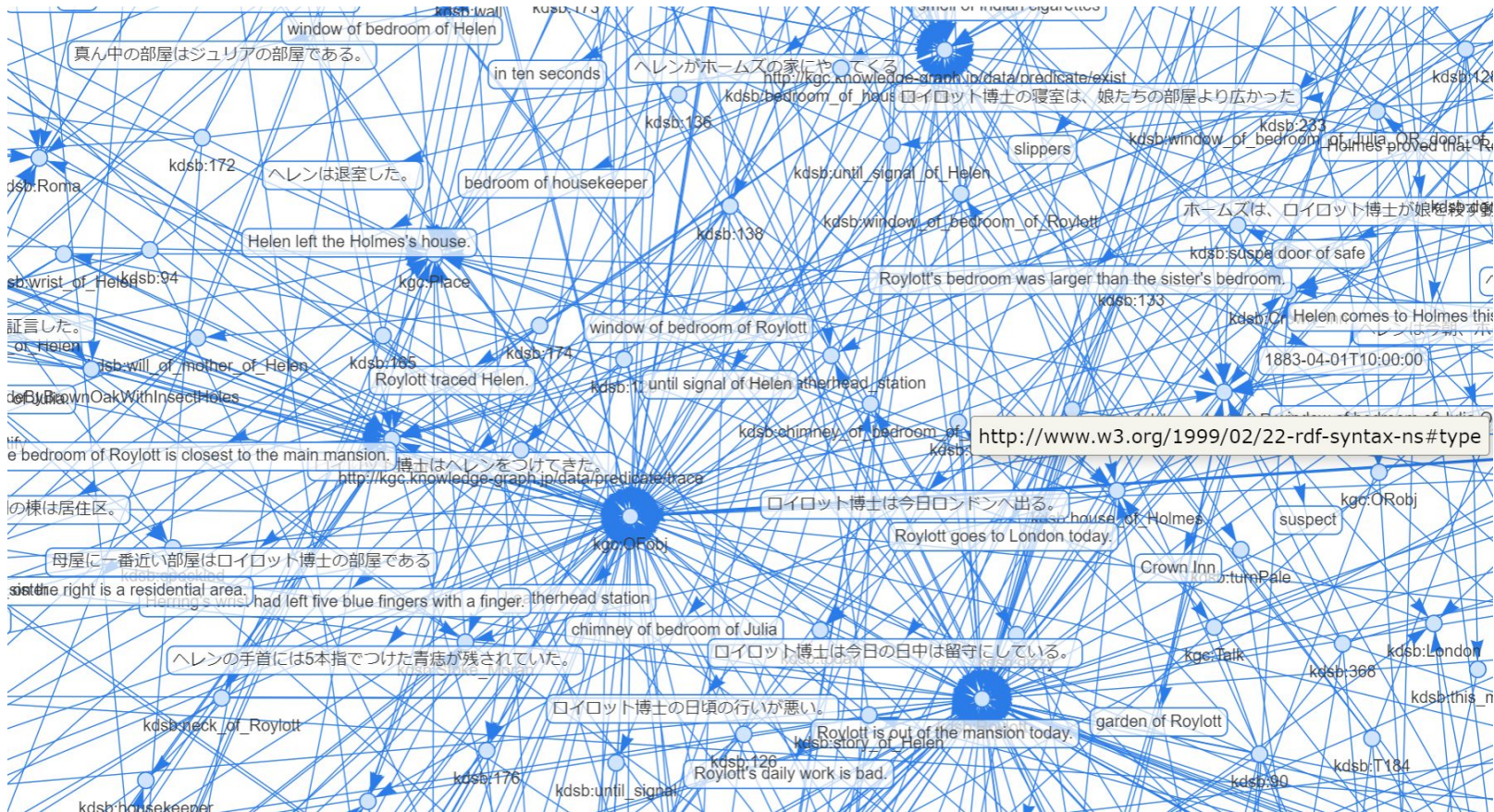


図2 従来の可視化ツール

可視化ツールを作成する作業手順

可視化ツールの作成は主に下の手順に分かれる。

1. SPARQLクエリの作成
1. Cytoscape.jsを用いた可視化ツールの作成

SPAQRLクエリの作成

- SPARQLクエリとは
 - ・RDFに対するクエリ言語
 - ・指定したグラフ構造に一致するトリプルを検索する
- 今回はシャーロックホームズという小説のSPAQRLクエリを用いて場面展開を中心とした可視化ツールを作成する。
- そのクエリ内で小説の原文と原文とのつながりの有るシーンと関係を示したものを取得して表示させる。各原文とシーンは日本語の物を取得した。


```

SELECT DISTINCT * FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
WHERE {
  ?s rdf:type/rdfs:subClassOf kgc:Scene .
  ?o rdf:type/rdfs:subClassOf kgc:Scene .
  ?s ?p ?o
  ?s kgc:source ?st.
  ?o kgc:source ?ot.
  FILTER(lang(?st)="ja") FILTER(lang(?ot)="ja")
}

```

図3 まだらの紐で使した SPARQLクエリ

	s	o	p	st	ot
1	<http://kgc.knowledge...	<http://kgc.knowledge...	kgc:then	"ヘレンは2ヶ月以内にお金を得る"@ja	"ヘレンはホームズにお礼する"@ja
2	<http://kgc.knowledge...	<http://kgc.knowledge...	kgc:when	"ロイロットはお金をもらう"@ja	"姉妹がロイロットと住んでいるとき"@ja
3	<http://kgc.knowledge...	<http://kgc.knowledge...	kgc:if	"お金は姉妹に分けられる"@ja	"姉妹が結婚すると"@ja

図4 まだらの紐のSPARQLクエリの結果

Cytoscape.jsを用いた可視化ツールの作成

- 文章のつながりを可視化するためにCytoscape.jsを使用した。nodeの作成とその関係をedgeで表現した。今回は場面展開が中心の分かりやすい可視化を行うのが目的である。
- そのため一目でわかりやすい場面の展開の可視化の図を作成し、詳細情報は画面遷移させ違うページで表示することが必要である。



図5 「まだらの紐」の場面間のつながりの取得結果

システムの結果

研究で得られた結果として原文のつながりとその詳細データへのページへ移動するという物の表示は出来るようになった。図5は全体図になっている。
公式が提供しているものよりも原文のつながりがなくないところはないところと離れていたり、何処の文章がつながりが多いのかが一目で分かるようになっている。

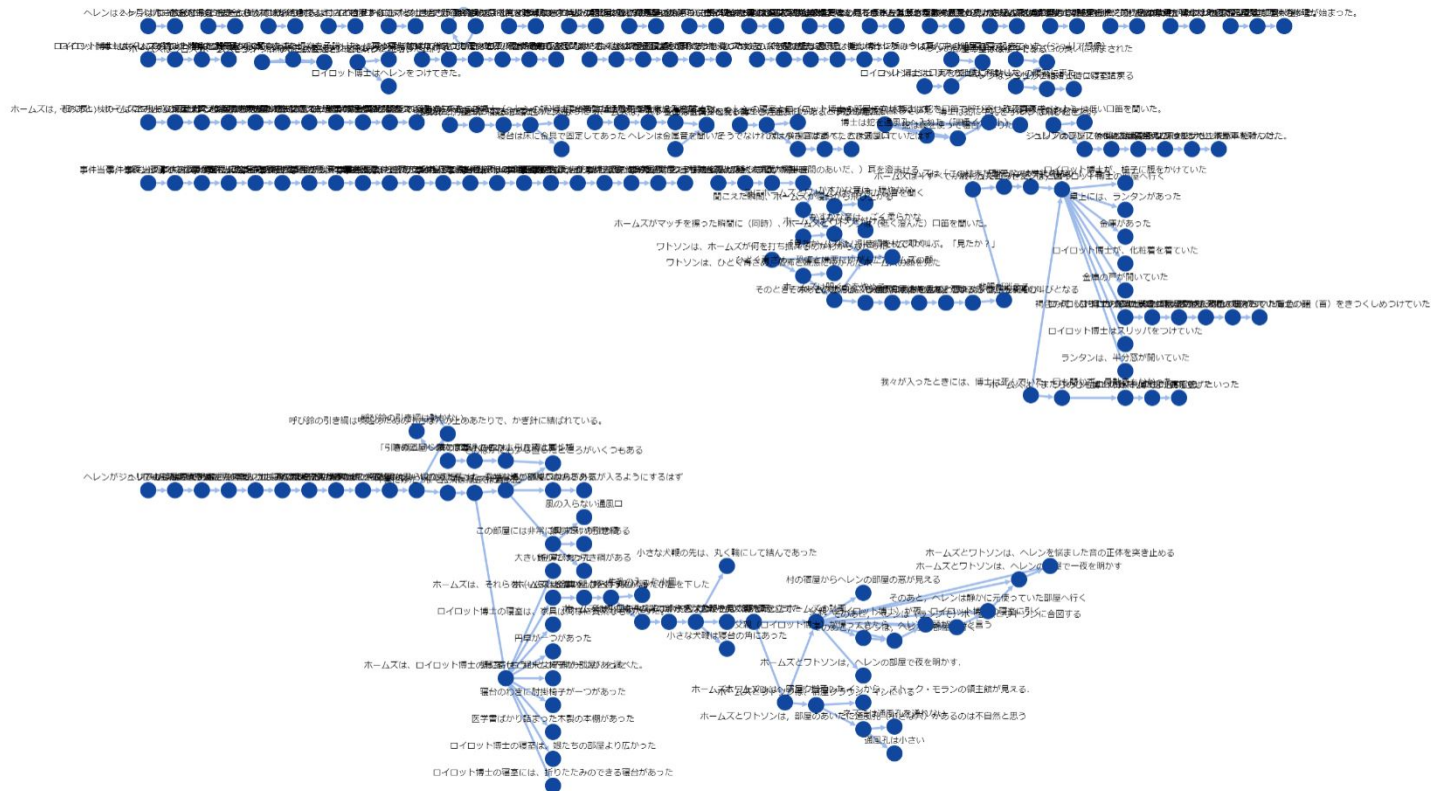


図6 「まだらの紐」の場面間の可視化結果の一部

まとめ

- 公式の提供している問題点のシーンとそのつながりのあるシーンの関係を見にくさはシーン同士を離したり、主語、述語、目的語を削除し、詳細ページで見るようにすることで改善された。しかし主語、述語、目的語のつながりを見れるようにすることでもっと分かりやすく出来ると思われる。
- なので今後の課題は主語、述語、目的語の可視化の追加をし、それを格納や展開できるようにしたり、見やすい詳細データの表示されているページの作成によってより一目でつながりが分かりやすくなり、ナレッジグラフの内容が把握しやすくなると考えられる。