



KGC: Knowledge Graph Completion Challenge

ナレッジグラフ推論チャレンジ

第1回ナレッジグラフ推論チャレンジ2018振り返り ～ まだらのひも *The Adventure of the Speckled Band* ～

第1回チャレンジ概要

- ➡ 対象小説 **まだらのひも** (Speckled Band) ※シャーロキアン人気No.1
- ➡ 応募件数 **8作品**(本部門:5, アイデア部門:3)
- ➡ 本部門(実装を伴う)
 - **【最優秀賞】** 野村総合研究所
 - III➡ **【優秀賞】** 上小田中411(富士通研究所)
 - III➡ **【ベストリソース賞】** FLL-ML(富士通研究所, 神戸常盤大学, 神戸市立西神戸医療センター)
 - teamOIF(立命館大学)
 - 塚越雄登(電気通信大学)
- ➡ アイデア部門(アイデアのみ, 実装を伴わない)
 - **【ベストアイデア賞】** 白松研 feat. 59(名古屋工業大学)
 - 生島高裕(数理先端技術研究所)
 - 橋本一成, 他(富士ゼロックス)
- ➡ 個々の詳細は公式サイト <http://challenge.knowledge-graph.jp> 参照

超簡単なあらすじ紹介（1／2）

➡ 依頼人ヘレンがホームズのところにやってくる

- ヘレンは最近、婚約した
- ヘレンはいま義父のロイロットと二人で暮らしている
- ロイロットは気性が荒く、近所でトラブルを起こしている
- ロイロットはインド帰りでヒヒやヒョウを飼っている
- ロイロットはロマと仲がよく、敷地を使わせている
- ロイロットが実母の遺産を管理している

➡ 2年前にはヘレンの姉ジュリアが変死していた

- ジュリアは婚約したばかりだった
- ジュリアはロイロットの隣の自室で死んだ
- 死ぬ前に「Speckled(まだらの)Band(楽隊？ひも？)」と言っていた

超簡単なあらすじ紹介（2／2）

- ➡ ホームズがロイロットの屋敷を訪れ、ヘレンの部屋で不寝番をする
 - ホームズは不審な物音を聞き、ムチで打ち付ける
 - 隣のロイロットの部屋から悲鳴が聞こえる
- ➡ ホームズがロイロットの部屋に入る
 - ロイロットの首に「まだらのひも(ヘビ)」が巻き付いていた
 - ロイロットはインドから取り寄せたヘビも飼っていた
 - インドのヘビの毒はイギリスではよく知られていなかった
 - ロイロットはヘビを飼いならしてジュリアを殺した(遺産分与を防ぐため)
 - ロイロットは同様にヘレンを殺そうとしたが、ホームズに阻まれた
 - ホームズに打ち据えられたヘビは、怒って飼い主ロイロットに噛み付いた

まだらのひものナレッジグラフ

➡ 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

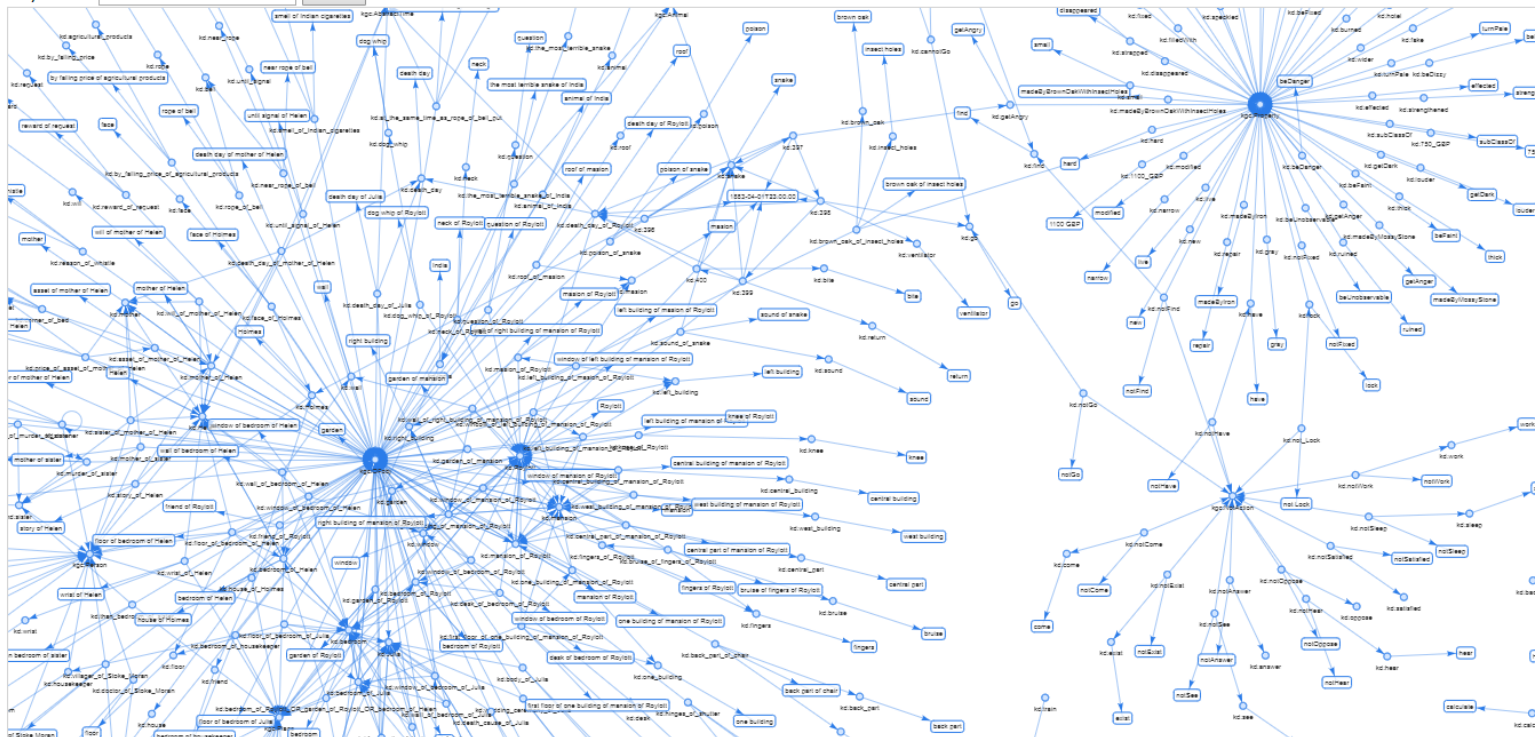
```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
PREFIX kd: <http://kgc.knowledge-graph.jp/data/SpeckledBand/>
SELECT ?s ?p ?o
WHERE {
  ?s ?p ?o .
  filter(?p != rdfs:comment)
  filter(?p != kgc:source)
  filter(?o != kgc:Situation)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

グラフDB(キーワード検索也可)

ナレッジグラフ(RDF形式)



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

8つの提案アプローチ

➤ 本部門（実装を伴う）

III ➡ 【最優秀賞】野村総合研究所

✓ 述語論理式を用いて定式化し、制約充足問題として解を導出

➤ 【優秀賞】上小田中411(富士通研究所)

➤ 【ベストリソース賞】FLL-ML(富士通研究所, 神戸常盤大学, 神戸市立西神戸医療センター)

➤ teamOIF(立命館大学) ☆学生

✓ オントロジー構築とSPARQL検索

➤ 塚越雄登(電気通信大学) ☆学生

✓ ルール推論とSPARQL検索

➤ アイデア部門（アイデアのみ、実装を伴わない）

III ➡ 【ベストアイデア賞】白松研 feat. 59(名古屋工業大学) ☆学生

✓ マルチエージェントによる議論モデル

➤ 生島高裕(数理先端技術研究所)

✓ 自然言語処理

➤ 橋本一成, 他(富士ゼロックス)

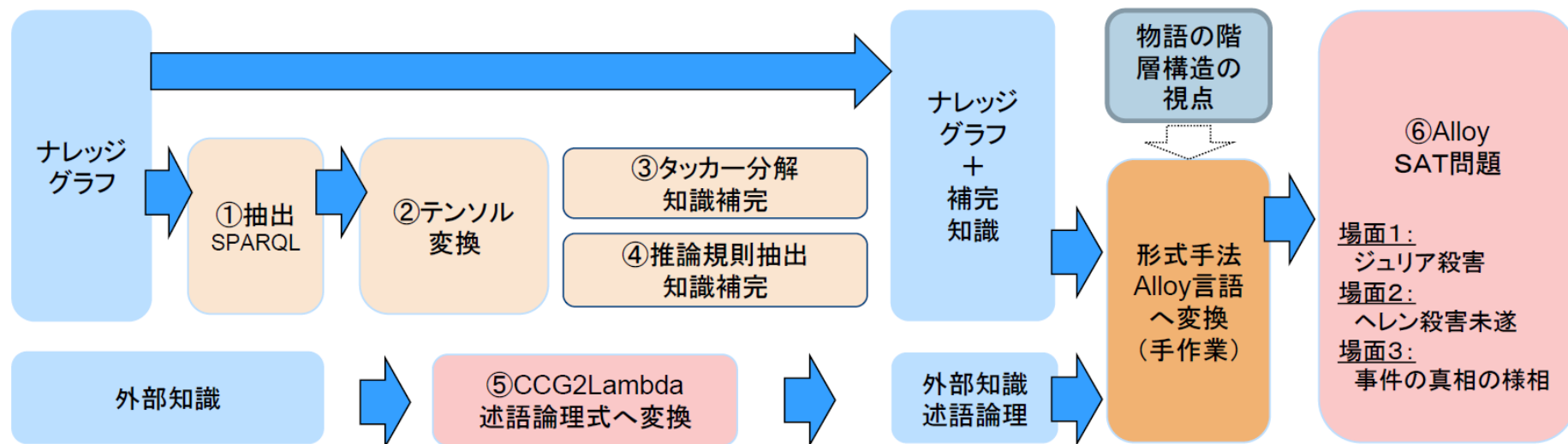
✓ イベントおよび推理に係る汎用的なオントロジーの構築

➤ 個々の詳細は公式サイト <http://challenge.knowledge-graph.jp/2018> 参照

【最優秀賞】野村総合研究所（1／4）

■アプローチの特徴

- 犯行状況を設定し、ナレッジグラフ（場面番号368番以下）＋外部知識を束縛条件に、述語論理式の充足可能問題として、真相を導く。



■結論

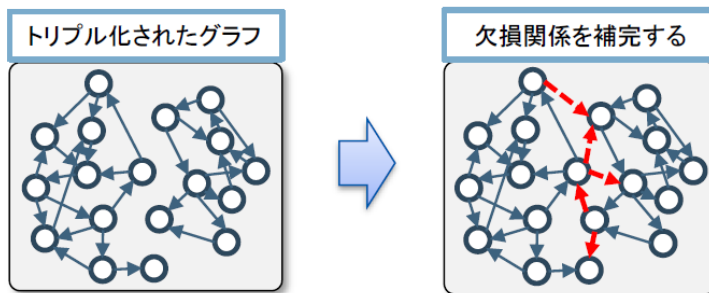
- 一連の事件（ジュリア殺害、ヘレン殺害未遂）は、ロイロットが犯人。
- 一部犯行は、ヘレンが犯人とも推察されるが、証拠不十分。

【最優秀賞】野村総合研究所（2／4）

③テンソルデータに対して、Tucker分解で、知識の補完

■与えられたナレッジグラフの全体傾向からリンクを補完する

- 外部知識を用いずに、「内部知識」でどこまで補完できるか。



全体の傾向から尤もらしいと考えられる関係を補完

※18(人物と動物) × 146(動作) × 18(人物と動物)のテンソル

■補完された知識(※場面は368番までの情報。)

- 行動や状況(部屋/時刻)が似ている人同士の間で、片方のみ行った行動が、他方の行動としても補完がなされた
- 補完された知識には解釈が困難なものもあるが、Juliaに対して、Animalが何らかの行為を行ったことが多数出る。

補完された知識	確度は低いが 補完候補になった知識
Homes See Animal	Helen bite Julia
Watson Hit Animal	Animal Call Julia
Watson Put Animal	Animal Meet Julia
Homes Go Roylott	Animal Support Roylott
etc	etc

※コアテンソルの大きさ(3 × 15 × 3)、繰り返し計算10数回程度で収束にいたった。

※疎テンソルのため、適用の妥当性については要検討。

【最優秀賞】野村総合研究所（3／4）

⑤ CCG2lambdaにて、外部知識(日本語文)を述語論理式に変換

■ 外部知識(日本語文)を、述語論理式に変換



例1:

日本文	全ての人は、自分を殺さない。
述語論理式	$\text{exists } x.(\text{exists } z1.(\text{全て}(z1) \ \& \ (x = z1)) \ \& \ \text{人}(x) \ \& \ \neg \text{exists } z1.(\text{自分}(z1) \ \& \ \text{exists } e.(\text{殺す}(e) \ \& \ (\text{Nom}(e) = x) \ \& \ (\text{Acc}(e) = z1))))$
Alloy言語形式	$\text{fact}\{\text{all } p:\text{Person} \text{all } t:\text{Time} \text{all } r:\text{Room}\{\text{no } p.\text{kill}.t.p.r\}\}$

例2:

日本文	ある人が、毒を盛られたならば、ある人は死ぬ。
述語論理式	$(\text{exists } x.(_ \text{人}(x) \ \& \ \text{exists } z2.(_ \text{毒}(z2) \ \& \ \text{exists } e.(_ \text{盛る}(e) \ \& \ \text{Past}(e) \ \& \ \text{exists } z2.(_ \text{人}(z2) \ \& \ (\text{Dat}(e) = z2)) \ \& \ (\text{Acc}(e) = x) \ \& \ (\text{Acc}(e) = z2)))) \rightarrow \text{exists } x.(_ \text{人}(x) \ \& \ \text{exists } e.(_ \text{死ぬ}(e) \ \& \ (\text{Nom}(e) = x))))$
Alloy言語形式	$\text{Fact}\{\text{all } q:\text{Person} \text{all } t:\text{Time}\{\text{some } \text{Person}.\text{毒を盛る}.t.q \Rightarrow \text{dead}[q,t.\text{next}]\}\}$

※述語論理式からAlloy言語に変換プログラムはできておらず、手作業で変換。

ただし、日本語文→述語論理式→CNF(連言標準形)→充足可能問題(SAT問題) という形ができるのが理想。

【最優秀賞】野村総合研究所（4／4）

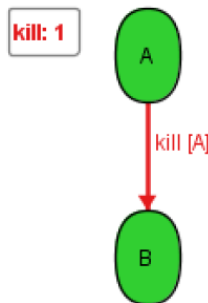
Alloyによる充足可能性の計算

- 「犯行の可能性の列挙」を、ナレッジグラフ+補完知識の述語論理式の充足可能性問題ととらえる。
- 充足可能性問題（以下、SAT問題）は、形式手法AlloyAnalyzerを使って解く。

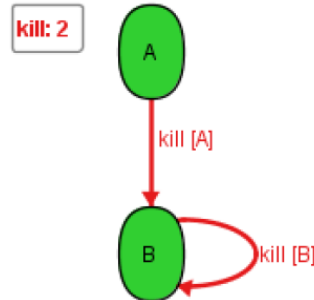
Alloyの例

構成	具体	Alloy形式言語
対象	ある人物A,Bがいる	<pre>abstract sig Person { kill : this -> Person } one sig A,B extends Person {}</pre>
論理	人は殺されると死ぬ	<pre>pred dead[q:Person] { some Person.kill.q }</pre>
事実	Aは生きている Bは死んでいる	<pre>fact{ not dead[A] dead[B] }</pre>

AがBを殺害する



AがBの自殺に協力する



Bが自身を殺害する



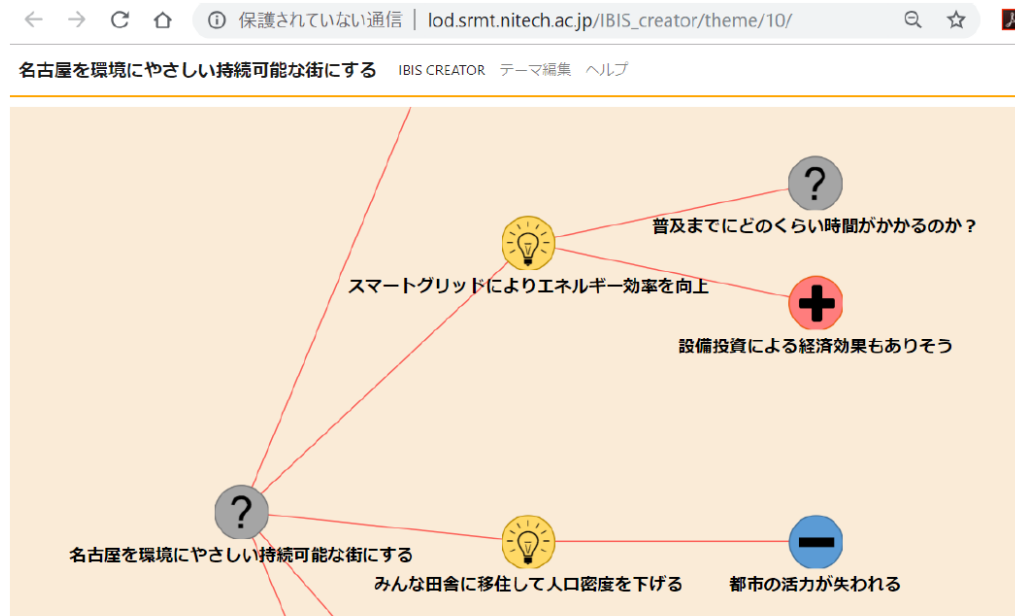
【ベストアイデア賞】名工大（1／4）

- 誰が犯人かを**複数エージェントで議論**することで、説明性を持つ推理AIを実現
 - 推理小説には、犯人を言い当てる探偵だけでなく、間違った推理を披露する助手役（サイドキック）も重要
 - それぞれの仮説を検証するような議論の場面を挿入することで、読者に事件背景や推理の根拠を理解させる働き
 - 議論内容をIBIS (Issue-based Information System) を用いて、課題、アイデア、アイデアの利点／欠点をノードとする木構造で表現[白松18]
- ✓ [白松+18] 白松俊, 池田雄人, 北川晃, 幸浦弘昂, 伊藤孝行: 自律的ファシリテータエージェントのための内容とプロセスを考慮した議論文脈理解モデルの検討, JSAI 2018, 1D2-OS-28a-03, 2018.
- 「犯人が誰か」という仮説をアイデアノードにすれば、IBIS構造は仮説の根拠を構造化するのに最適なのでは？

【ベストアイデア賞】名工大（2／4）

IBIS構造を扱うツール

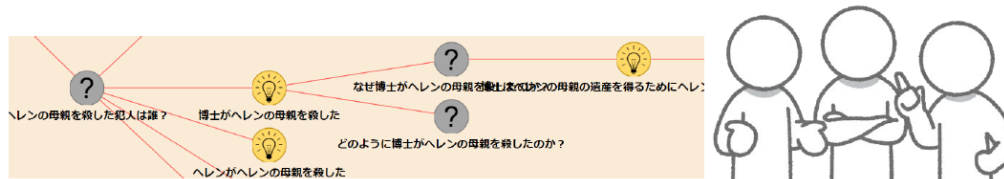
- IBIS CREATOR http://lod.srmt.nitech.ac.jp/IBIS_creator/
 - IBIS構造を作成・編集できるWebアプリ（白松研 神谷が開発）
 - 作ったIBIS構想はLinked Open Data化される
 - 簡単なWeb APIを備え，外部システムから操作可能



【ベストアイデア賞】名工大（3／4）

推理のための議論手順の概要

- 「誰が犯人か」という仮説を担当する議論エージェント
 - ナレッジグラフ中の登場人物 x ごとに「 x が犯人」という仮説を立て、その仮説の説明を試みる議論エージェントを割り当てる
- 仮説の詳細を質問するファシリテータエージェント
 - 「 x はどのように殺したのか？」のような質問をIBIS構造に追加
- 各議論エージェントはその質問の答えとなる仮説を生成
 - 生成された仮説は、IBIS構造に追加されると同時に、
 - 各議論エージェントが保持するナレッジグラフを補完

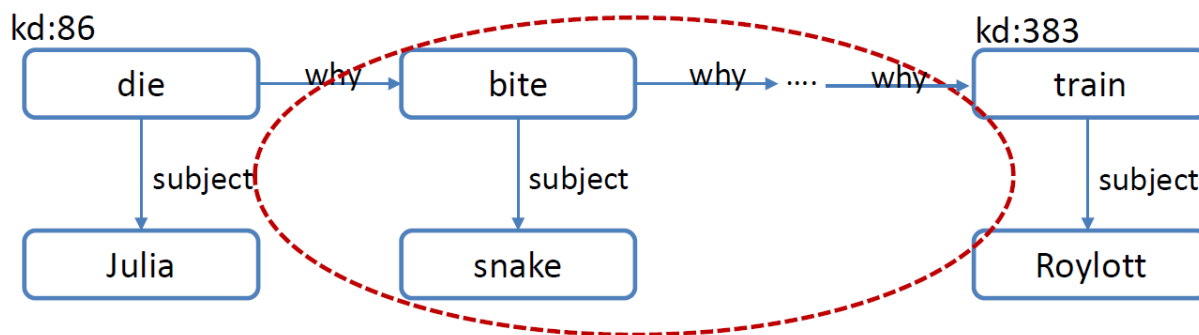


- ファシリテータエージェントが説明の整合性を評価
 - 各議論エージェントが生成した説明の整合性を評価し、最終的に評価の高かった説明を出力

【ベストアイデア賞】名工大（4／4）

想定する仮説生成例

- 例えば「博士はどのようにジュリアを殺したか？」に関する仮説は、以下のようなkgc:whyの連鎖の構造で生成



- このような仮説でナレッジグラフが補完されていれば、以下のようなSPARQLクエリで犯人 x と殺人行為 $cause_id$ を検索可能

```
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 distinct ?x ?cause_id WHERE {
  ?die_id kgc:hasPredicate kd:die.
  ?die_id kgc:subject kd:Julia.
  ?die_id kgc:why+ ?cause_id.
  ?cause_id kgc:subject ?x.
  ?x rdfs:type kgc:Person.
}
```

評価方針

- 説明性を有する推論・推定技術を評価するためには、適切な指標設計が必要
 - 正解を導いていることに加えて、説明性、実用性、新規性、処理時間など、複数の指標を設計し、各アプローチのメリットやデメリットを客観的に評価、分類、体系化
- 一方で、定量的評価に加えて、評価者側と応募者側との議論(ピアレビュー)を通じた定性的な比較や評価、問題意識の共有も必要
- DARPA XAIでは、現在の機械学習技術においては精度と説明性にトレードオフがあることを明言しており、両者を測定するとしている。
- 更に、説明性は心理学的にアプローチされており、説明の効果測定ではその明快さや有用性に関するユーザーの満足度をレーティングする
- これらを参考に、**第1回チャレンジでは、基本情報を共有した後、専門家評価、一般評価の組み合わせで評価を実施した**

基本情報の共有

- ▶ 主催者側で以下の基本情報を調査し、事前に審査員である専門家と共有（今回、審査員はSWO研究会の幹事を中心に7名に委託）
 - 正解が出せているかどうか？
方法の如何にはよらないが、推論・推定した犯人は正しいか？
ここで犯人とは、小説の中で犯人されている人物とする。
 - プログラムが実行できるか？
提出されたプログラムが正しく動作し、結果が再現できたか？（アイデア部門を除く）
但し、全ての動作の再現を保証するわけではない。
 - プログラムのパフォーマンス
プログラムの動作環境、動作速度に関する何らかの情報
 - 使用した知識・データ量
提供したナレッジグラフの内、ID 何番までを使用したか？
外部知識・データを活用している場合は、その情報。

専門家による評価

▶ 専門家評価では、以下の項目について審査員が5段階評価

▶ 推論・推定方法

- 技術性(Significance)
推論・推定方法の技術的工夫。
- 汎用性(Applicability)
他の問題にも適用できる手法であるか？
目安としては、
3:他の推理小説にも適用可能
5:他のドメインにも適用可能)
- 発展性(Expansibility)
今後の技術的な発展が期待できるか？
例えば、現状の問題点を解決すると改善が見込める、など

専門家による評価

➡ 知識・データ

➤ 知識・データ構築の工夫

知識/学習データを構築にかけた工夫(量×質×プロセス). 例えば, 外部知識・データをどれだけ用意したや, 外部知識・データの作り方の工夫など

➤ 知識・データ活用の工夫

提供されたナレッジグラフや自身で構築した知識を効果的に利用しているか? 少量の外部知識でスマートに解いたや, 大量の知識で少ステップで解いた, など

➡ その他

➤ 実現可能性(アイデア部門のみ)

技術, 知識・データ構築/活用の双方を含めたアイデアの実現可能性.

➤ 論理的説明性

論理的な説明が成立しているか?

1: 説明や根拠の提示が全くない.

3: 根拠となるエビデンスが何らかの形で提示されている.

5: 必要十分な推論(推定)過程において, 一貫性を持った説明がなされている.

➤ 努力性 作品(知識/データ/システム)作成にかけた労力

➤ 総合的なコメント(自由記述)

一般審査

- ➡ 専門家が論理的に説明が通るかどうかを判断したのに対し、一般審査では時間的な制約から説明の心理的な側面(納得性)に絞って評価を実施
- ➡ 各応募者が本部門15分, アイデア部門10分で応募作品の説明を行い, 参加者45名が以下の2点についてアンケートに回答
 - 総合評価
アイデアの面白さ, プレゼンの良さなども含めた総合的な評価。
 - 説明性
「あなたが裁判員(陪審員)だったとして, その説明に納得にできますか？」
- ➡ 総合評価を設けたのは, プレゼンの上手さや面白さなど説明性以外の評価を総合評価に回してもらうため
- ➡ 説明性のみの評価では, それら別の観点の評価も混じってしまうと考え, あえて他の観点を含む項目を設けた。

評価結果

専門家による指標毎の審査結果(平均)

応募番号	A	B	C	D	E	F	G	H
技術性	3.43	3.29	4.14	3.00	4.71	2.29	3.43	3.43
汎用性	2.86	3.50	3.00	2.00	3.71	2.33	3.86	3.86
発展性	3.14	3.67	4.00	2.14	4.14	3.17	3.29	3.00
知識・データ構築の工夫	3.14	4.00	4.14	2.71	4.14	2.33	3.29	2.57
知識・データ活用の工夫	3.43	3.17	4.14	2.86	4.14	1.83	3.14	2.57
実現可能性（アイデアのみ）						2.17	3.57	2.86
論理的説明性	3.57	2.33	4.71	3.43	4.00	2.50	3.43	3.00
努力性	3.57	4.17	4.57	3.00	4.71	3.17	4.00	3.00
平均値	3.31	3.45	4.10	2.73	4.22	2.52	3.49	3.06

専門家および一般による審査結果

評価指標	総合評価		説明性	
	1 位	2 位	1 位	2 位
平均値（一般）	4.26	4.04	4.11	3.97
中央値（一般）	4.50	4.00	4.00	4.00
標準偏差（一般）	0.67	0.65	0.78	0.69
平均値（専門家）	4.22	4.10	4.00	4.71
中央値（専門家）	—	—	4.00	5.00

評価結果

- 一般審査の結果は、総合評価と説明性の平均値、中央値、標準偏差を比較
- 1位と2位を比較では、
 - 平均値では総合評価と説明性いずれも1位が2位を上回っているが、
 - 中央値では総合評価は上回っているが説明性では同値となり、
 - 標準偏差では総合評価と説明性いずれも1位のほうが大きかった。
- t 検定($p = 0.05$)では、総合評価では有意差が得られたが説明性では得られなかった
- 専門家評価で1位と2位を比較すると、
 - 各指標の平均では1位が2位を上回っているが、
 - 説明性の観点では結果が逆転した(t 検定による有意差も認められた)
- 尚、いずれの指標でも標準偏差(全作品平均)は <1.0 、大きな意見の相違はなかった
 - ばらつきが最も少なかったのが論理的説明性であり、
 - ばらつきが最も大きかったのは努力性であった...
- 結果として、1位と2位の順位付けは専門家審査員による議論に任せられたが、説明性以外ではいずれも1位は2位と同値または上回っていたことから本結果とした
- 同時に、本チャレンジの一番の課題である説明込みでの評価については課題を残した
- その他に各専門家のコメント等を踏まえてベストリソース賞とアイデア賞を授与

参考文献

- Report on the First Knowledge Graph Reasoning Challenge 2018 - Toward the eXplainable AI System, Corr abs/1908.08184, 2019
- **第1回ナレッジグラフ推論チャレンジ2018開催報告～ 第2回チャレンジ開催案内～, 第33回人工知能学会全国大会, 1K2-J-4-04, 2019**
- **第1回ナレッジグラフ推論チャレンジ2018開催報告 — 説明性のある人工知能システムを目指して—, 人工知能学会学会誌34巻3号, pp.396-412, 2019**
- **その他の参考資料は以下をご参照ください**
 - <https://challenge.knowledge-graph.jp/reference.html>