Contextualized Scene Knowledge Graphs for XAI Benchmarking

Takahiro Kawamura
kgrc@knowledge-graph.jp
National Agriculture and Food Research Organization
Japan

Shusaku Egami
National Institute of Advanced Industrial Science and Technology
Japan

Kyoumoto Matsushita
Fujitsu Limited
Japan

Takanori Ugai
Fujitsu Limited
Japan

Ken Fukuda
National Institute of Advanced Industrial Science and Technology
Japan

Kouji Kozaki
Osaka Electro-Communication University
Japan

ABSTRACT

In order to utilize artificial intelligence (AI) safely and securely in society, explainable artificial intelligence (XAI) technology, which has the property of being able to explain the reasons why a system has reached a conclusion, is necessary. Therefore, although machine learning approaches are currently the mainstream of AI, AI technology that combines inductive machine learning and deductive knowledge utilization is expected to become necessary in the future. Currently, however, there is no dataset to evaluate both approaches properly. In this study, we constructed and refined large-scale scene graphs and event-centered knowledge graphs, and have released them as open data. While most knowledge graphs contain only simple relationships, the constructed knowledge graphs are characterized by the fact that they contain more complex relationships that reflect the real world, such as temporal, causal, and probabilistic relationships. In addition, we developed refinement methods for the actual use of the constructed knowledge graphs for inference and machine learning. We held four technical competitions in Japan for AI technologies with various explanatory possibilities, gathered methods related to inference and estimation from a wide range of IT engineers, and classified the proposed technologies. An international version of the competition is planned for FY2022. In the future, we would like to design appropriate indices and conduct objective evaluations, classifications, and systematization for the development of AI technologies with explanatory properties, especially those that combine inductive machine learning (inference) and deductive knowledge utilization (reasoning).

CCS CONCEPTS

• **Theory of computation → Semantics and reasoning**; **Logic**;
• **Computing methodologies → Artificial intelligence**.

KEYWORDS

Event-centric, Knowledge Representation, Linked Data, Open Data

IJCKG2022, Oct. 27-29, 2022, Hybrid

ACM Reference Format:
Takahiro Kawamura, Shusaku Egami, Kyoumoto Matsushita, Takanori Ugai, Ken Fukuda, and Kouji Kozaki. 2022. Contextualized Scene Knowledge Graphs for XAI Benchmarking. In *IJCKG2022: The 11th International Joint Conference on Knowledge Graphs.* Online, 9 pages.

1 INTRODUCTION

In recent years, there has been a growing interest in artificial intelligence (AI) technologies such as deep learning. In the near future, AI technology is expected to be incorporated into various social systems. It is also expected that AI technology will be integrated into various social systems, and such systems will eventually leave human hands alone to make decisive decisions on their own. However, in order to utilize AI in society in a safe and secure manner, technologies and quality assurance are needed to confirm that AI systems are operating properly. However, machine learning technologies such as deep learning are black boxes in their decision-making and prediction processes, and humans do not understand the reasons for their conclusions. For this reason, AI technologies that have explainability (i.e., XAI) or interpretability, properties that enable an AI system to explain the reasons for its conclusions, are attracting attention. Indeed, this has already been recognized domestically and internationally, and related workshops and symposiums have been actively held at international conferences on deep learning.

In addition, although current AI technologies are mainly based on machine learning approaches, AI technologies that combine inductive machine learning and deductive knowledge utilization are expected to become necessary in the future. For example, when considering automated driving as an AI-applied system that cannot escape accountability in the event of an incident or accident, it is essential to recognize the situation around one's own vehicle using machine learning and estimation techniques based on sensor data obtained from the vehicle's front camera and radar. However, traffic rules and vehicle operation itself are predefined knowledge, and automated driving is ultimately based on the integration and fusion of knowledge and data.

Currently, however, there are no datasets that can be used to evaluate inductive machine learning techniques and deductive knowledge application techniques properly. Most benchmark datasets used for estimating relationships by machine learning contain only relatively simple relationships and cannot be used for problems that combine multiple subtasks to achieve an overall goal, such as automated driving. For example, representative datasets such as FB15k and WN18 contain only a large amount of simple relationships between data points, such as "is-a" and "hasSpouse". On the other hand, much of the rule-like knowledge is domain-dependent, and very few datasets can be used as large, generic test sets that are also applicable to machine learning. A suitable dataset for the task of implementing interpretable AI using inference and estimation techniques should include not only relatively simple relationships, such as those for which estimation of binary relations suffices, but also more complex relationships that reflect the real world, for example, temporal, causal, and probabilistic relationships.

Therefore, in this study:

(1) As a common dataset for the evaluation of inference and estimation techniques that satisfy the above requirements, we constructed a scene- or event-centered knowledge graphs that contain complex and structural relationships, such as real social problems and human relationships, and created guidelines for their refinement.
(2) We have published these graphs as open data and held technology competitions four times in the past[5, 6, 8, 10, 14, 15] to gather methods related to inference and estimation from a wide range of IT engineers and researchers, and classified the proposed technologies.

Our goal is to design appropriate indicators and then objectively evaluate, classify, and systematize AI technologies with explanatory and interpretative properties, especially those that combine inductive machine learning (estimation) and deductive knowledge (inference) applications.

The schema of a knowledge graph involves dividing a series of contents into the smallest units (scenes), and to graph the contents of each scene and the relationship between scenes. Such a knowledge graph is called a scene knowledge graph or an event-centric knowledge graph[2, 4]. Since a scene consists of a large number of geographic objects and their spatial relations and attributes, it is important for geographic information system (GIS) applications to explore effective methods to organize spatial scenes so that they are more readable by humans and machines. For example, such methods are being used in practical real-world applications, such as automatic driving, as shown earlier. These methods are also widely used in simulation games and the metaverse, as well as in the prediction of dangerous behaviors for the safety and security of the elderly[1].

In the following sections of this paper, Section 2 lists the related works on event-centered knowledge graphs, and Section 3 describes the schema design and construction procedure of the knowledge graphs of this paper. In Section 4, we present a guideline for the refinement of the constructed knowledge graphs and describe the verification results on the applicability of the guidelines. In Section 5, we present an overview of the technical challenges we have faced so far. However, this paper focuses on the construction of knowledge graphs, and the details of the proposed techniques for the technical challenges are left for another paper. Finally, in Section 6, we summarize the results and discuss future challenges.

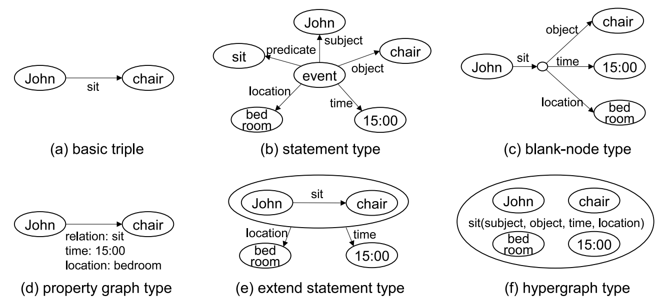

Figure 1: Schema for event and scene graphs.

2 RELATED WORKS

Knowledge graphs can be used to describe static relationships between things, such as in product data, a thesaurus, and human relationships, as well as events that occur in space and time, such as observational data. In recent years, knowledge graphing of events or scenes, such as video content[9, 17], has been actively studied. Several schema patterns have been proposed for knowledge graphing of events or scenes, as shown in Figure 1. There seems to be no significant difference in the difficulty of construction in either case. Figure 1(a) is a basic triple pattern for knowledge graphing the information "John sits in a chair". If we consider representing this information as an event by adding time and a location to it, we can take a structuring method such as shown in Figures 1(b-–e). In Figure 1(b), all information is linked around an event node, such as with the simple event model (SEM)[16]. Figure 1(c) shows a pattern that supplements the status of the predicate by inserting a blank node before the object, and is positioned as an alternative representation of the SEM. Figure 1(d) displays a pattern in which the edges hold multiple pieces of information. It is generally called a property graph and is supported by graph database systems such as Neo4j. In Figure 1(e), meta-information is given to the entire triple, and standardization is being discussed in the form of a resource description framework (RDF)-star. Figure 1(f) portrays a knowledge hypergraph that can express relations other than binary relations, and its application to Earth observation data has been demonstrated[12]. Thus, different schema patterns can be applied to knowledge graphs of event information.

In contrast, this study aims to annotate the events themselves, such as "events with a risk of accidents," and to determine such events automatically using AI technology. Therefore, a node representing the event itself is necessary, and the statement type shown in Figure 1(b) is adopted. Details of the schema are given in Section 3.1.

On the other hand, the contents in the event or scene knowledge graphs include EventKG[3], ECKG[13], Drammer[11], and so on. EventKG is a knowledge graph describing 690,000 contemporary and historical events and incidents for the purpose of answering questions and generating histories (timelines) from a specific perspective. The schema is based on the aforementioned SEM[16] and is extended to express temporal relationships, and so forth. It has many similarities with our schema, such as definitions of relationships among events. However, the granularity of its target events is considerably larger than in our scenes, and it is difficult to represent information such as who, when, and how for each scene using EventKG's model (although it is possible to describe them, it would be a complex graph, and it would be difficult to construct and search the dataset). ECKG provides its own model to annotate information extracted when building a knowledge graph directly from news events written in a natural language. It provides a unique model. However, the model is simple (only who, what, where, and when) because automatic extraction is the subject matter. Drammer is not simply a chronological representation and comparison of narrative content, but is fiction–specific. It is an ontology that includes conflicts between characters, segmentations of the narrative, and definitions of emotion and belief for more dramatic representations. It was constructed by analyzing many dramas, but its purpose is different from that of this study, which is intended to represent facts (including falsehoods) in the real world.

By contrast, in this study we (1) constructed knowledge graphs that convert the background of the case and the characters into knowledge, using a mystery novel as a subject, and (2) conducted a technical challenge to identify correctly the culprit and cause of a case or an accident from given information using inference and estimation techniques to explain the reasons (evidence, tricks, etc.) for such identification appropriately. The reasons for choosing mystery stories as the subject matter include:

- They can allow for the design of tasks that are virtually closed (e.g., which have answers and can control the constraints that lead to them) while including complex relationships in the real world.
- Some tasks can be solved without including probabilistic processing or machine learning, such as uncertain information or photographic evidence, or without supplementing common knowledge that is not written explicitly, thus encouraging the fusion of estimation and inference.
- They have an explanatory quality to human beings that the reader must agree with in order for it to work as a novel.
- The stories are widely known to the public and easily attract interest.

As for this technical challenge, in top conferences of AI and neural networks, such as IJCAI, AAAI, NIPS, and ICML, papers and workshops that have "explainability" as a keyword and that analyze the properties of AI models have significantly increased since 2016. However, no other research activity exists like the challenge discussed in this work, which uses knowledge graphs, including social problems as common test sets, and tries to solve the problems with explainability (i.e., using XAI), aiming to integrate inductive estimation and deductive reasoning.

3 KNOWLEDGE GRAPH CONSTRUCTION

In this study, the contents of eight of Sherlock Holmes's short mystery stories, "The Speckled Band", "The Dancing Men", "A Case Of Identity", "The Devil's Foot", "The Crooked Man", "The Abbey Grange", "The Resident Patient", and "Silver Blaze", were converted into knowledge graphs based on events and scenes. Participants in the technical challenge developed AI systems using these data together with their own external knowledge, which was added

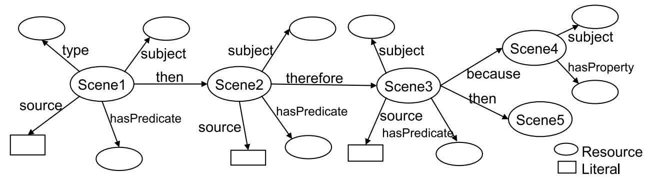

Figure 2: Scene knowledge graph.

and created as necessary. The knowledge graph and past proposed techniques are available on the official website[1].

3.1 Knowledge graph schema

In designing the knowledge graph schema, five open workshops were held in Japan in 2017–2018 to discuss the basic and detailed design of the schema, knowledge graph construction methods, and concrete construction work. The total number of participants was about 110. In the workshop, we first examined the knowledge required for inference and estimation (what should be described in the knowledge graph) and how it should be expressed through a pilot description of the knowledge graph. Then, based on the feedback from the participants, we decided on a basic policy of describing the people, things, and places involved in each scene, focusing on the scenes depicted in the scenes and the relationships between the scenes. When designing the schema, in addition to expressiveness to represent the subject novels, we also considered the ease of constructing the knowledge graph and the convenience of providing it as data for inference processing, and decided on a schema with mainly 5W1H edges, focusing on scenes (Figure 1(b)). Thus, a mystery story is represented by each scene and the relationships among scenes. Each scene[2] in a mystery story is assigned a unique internationalized resource identifier (IRI), which is used as the subject to describe a scene in the story by adding information about people, organizations, and places as objects. The relationships between scenes explain the causal relationships of chronological actions and events by referring to the IRIs. This is how a series of storylines is expressed. In addition, rules and table data can be linked to describe common sense data such as axioms and to represent information such as timetables. The content of the story is stored as literal values for natural language processing. Figure 2 shows an example knowledge graph.

The following basic properties are provided for describing each scene. In order to summarize the information associated with a scene, this property takes the scene as its subject. Note that it is not in the general <subject, predicate, object> format. Figure 3 shows an example scene description.

- subject: a person or thing that is the subject in the description of the scene.
- hasPredicate: a predicate that describes the content of the scene.
- hasProperty: a property of the person or thing that is the subject of the scene description.

[1]https://challenge.knowledge-graph.jp/
[2]Only scenes that are judged to be necessary for the deduction are converted to knowledge graphs.

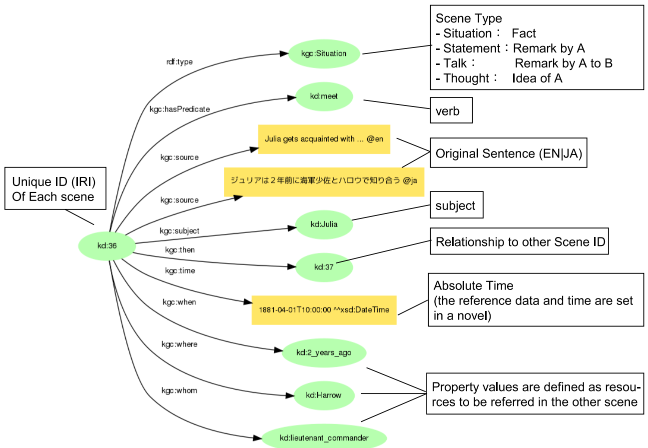

Figure 3: Example of a scene

- Objects that describe the details of the scene: who/whom, where, when, what, how, etc.
- Relationships between scenes: then, if, because, etc.
- time: absolute time when the scene occurred (xsd:DateTime).
- source: original text of the scene (English/Japanese; literal).

In order to distinguish whether each scene is a fact, someone's assertion (statement), or someone's idea, they are classified into the four types shown in Table 1.

Table 1: Scene types

Scene type	Description
Situation	facts and circumstances, e.g., John was murdered.
Statement	assertion by someone, e.g., Taro is lying in "John says Taro is lying."
Thought	someone's idea, e.g., Taro is the culprit in "John thinks Taro is the culprit."
Talk	John's remarks to Taro

Typical predicates representing the relationships between scenes are shown in Table 2.

Table 2: Relationships between scenes

Relation type	Predicate
Time	then, at the same time, when, after, before
Condition	if
Reason	because, why

A list of classes and properties is shown in Figure 4.

In order to express AND OR relations when there are multiple subjects and objects, the AND relation is expressed by describing multiple triples with the same property. The OR relation is described by describing resources that represent OR combinations as instances of the ORobj type. From this resource, multiple resources that are the target of OR are described via the kgcc:orTarget property. In addition, to handle the negation of predicates (not, cannot), we introduce classes of the notAction and cannotAction types

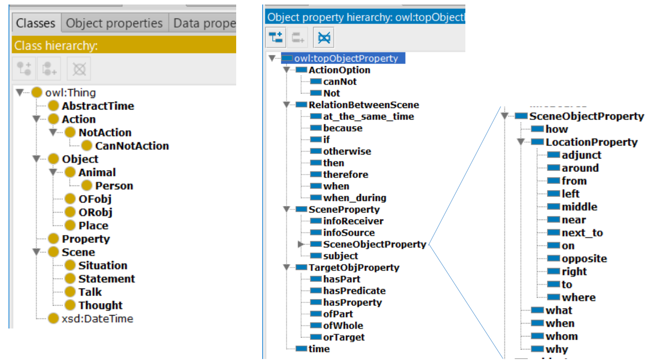

Figure 4: Class and Property lists.

as subclasses of Action. Negative predicates are described as instances of these classes. At the same time, they are connected to predicates of the affirmative form (Action type) by the kgcc:Not and kgcc:canNot properties.

3.2 Procedure of knowledge graph construction

The following procedure was used to convert the eight mystery stories into knowledge graphs:

(1) Extract sentences necessary for deduction from mystery stories (in Japanese) whose copyrights have expired.
For each novel, about 300 to 500 sentences were extracted. Since arbitrariness enters into the extraction of the parts necessary for deduction, it is better to include information that is not directly related to the identification of the murderer (scenery, description of the situation, common sense, etc.) in the knowledge graph. However, some extraction is currently performed to reduce the number of man-hours involved.
(2) Rewriting the original text into sentences with clear a subject and object (i.e., short sentences).
One short sentence corresponds to one scene on the knowledge graph.
(3) Assign semantic roles (e.g., 5W1H) to phrases using natural language processing tools.
Japanese semantic roll labeling technology is used for semantic role assignment. The results are output as a predicate and an object for each scene in a spreadsheet, and are visually checked at the end.
(4) Control orthographical variants.
We eliminate any notational distortions on a novel-by-novel basis and across novels as much as possible during the construction phase. Further refinement is shown in Section 4.
(5) Add relationships between scenes (e.g., temporal relationships).
(6) Translate the source text into English and convert the entire text into a knowledge graph.

An example of the application of steps (2)–(6) is shown in Figure 5. Note that the series of tasks were performed by part time students and software engineers (general programmers, not advanced knowledge engineers). The costs for knowledge graph construction of each story are as follows: (1) 3 hours per person, (2) 20 hours per person, (3) 5 hours per person, (4) 7 hours per person, (5) 3 hours per person, and (6) about 1 hour per person.

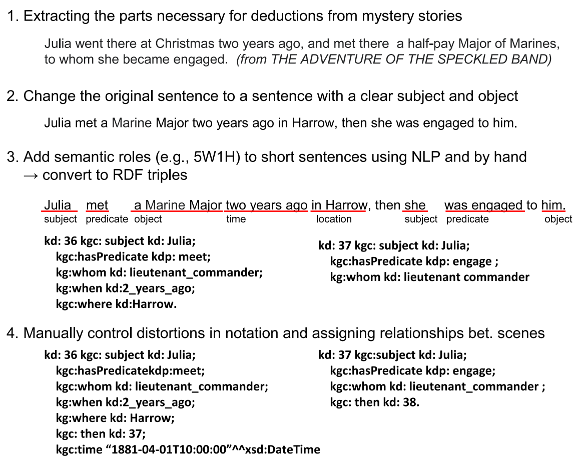

Figure 5: Example of knowledge graph conversion.

4 GUIDELINES FOR KNOWLEDGE GRAPH REFINEMENT

In this study, we have made a series of improvements to the content of the descriptions in a knowledge graph in order to identify the culprit and find the culprit's motive in mystery stories, expressed as a knowledge graph. In the following, we present a guideline consisting of ten items/steps we found through the refinement process of the knowledge graphs. Finally, to validate the applicability of the guidelines, findings obtained through the implementation of the guidelines by a third party are presented. Steps 4.2-–4.3 are related to the addition of implicit information that is not explicitly described. Steps 4.1 and 4.4-–4.7 are related to the unification of triple structure, and steps 4.8-–10 are related to the unification of the vocabulary.

4.1 Short sentences in English are converted to a syntax that is easy to change to an RDF

- Clarify the division between subject and predicate.
  For example, in the sentence "There is no place for Percy Trivellian," it is difficult to tell whether the subject is "Percy Trivellian" or "Percy Trivellian's place," and also whether the predicate is "does not have place" or "does not exist". In such sentences, since the target to which information is to be added is "Percy Trivellian," the subject is taken to be "Percy Trivellian".
- Complements omitted objects, complements, places, and so forth.
  - Example 1: although there is no location information in "Helen lives with Roylott," it is clear from the context that she lives in "Roylott's house," so the place should be added.
  - Example 2: in the phrase "Roylott is a father-in-law", it is not clear from whose perspective he is a father-in-law, so the additional information should be provided.

4.2 Adding implicit scenes

For example, "the day Helen's mother died" can be expressed as a single literal as "death_day_of_mother_of_helen". However, it cannot be used for inference because it does not logically express the information that this is the day that Helen's mother died. Therefore, we introduce a new scene in which "Helen"'s "mother" "died".

4.3 Add time information

If there is no description of time in the text, absolute time is given to each scene to the extent that it does not affect the narrative. As such, "then," "before," "after," and so forth are added as connections between scenes to clarify the chronological information.

4.4 Screening of sentences to be treated as scenes

For example, the sentence "The money is 1000 pounds a year" cannot be understood as a stand-alone scene. Therefore, a triple is added to supplement the scene such as "Helen and Julia receive their inheritance."

4.5 Unification of triplication from typical sentence patterns

For example, "there is" and "exists" are unified into "exists" to standardize the symbols used in the inference process. Also, information (adjectives) that describe properties are unified with the value hasProperty;
e.g., "Mr. A's salary" hasProperty [value 100, unit: dollar].

4.6 Division when there is more than one subject or object

For example, the scene "Holmes and Watson got out of the carriage" splits the subject (value of the subject) into two parts, "Holmes" and "Watson". Also, the scene "Holmes placed a box of matches and a burnt candle near a long, thin walking stick" splits the object (value of kgc:what) into "a box of matches" and "a candle".

4.7 Typing at nesting

In order to express appropriately nested structures caused by hearsay, each utterance is decomposed as a scene and given an appropriate type and source of information. For example, the scene "Holmes said that "Mr. B said that "Mr. A said (any)""." is decomposed as follows:

```
# Holmes said kd:id-x
kd:id-a rdf:type kgc:Situation ;
    kgc:subject kd:Holmes ;
    kgc:hasPredicate kdp:say ;
    kgc:what kd:id-x .
# Mr. A said kd:id-y
kd:id-x rdf:type kgc:Statement ;
    kgc:InfoSource kd:Holmes ;
    kgc:subject kd:A ;
    kgc:hasPredicate kdp:say ;
    kgc:what kd:id-y .
# Mr. B said (any)
kd:id-y rdf:type kgc:Statement ;
```

```
kgc:InfoSource kd:A ;
kgc:subject kd:B ;
kgc:hasPredicate kdp:say ;
kgc:what (any) .
```

Then, we distinguish whether the scene type is Situation or Statement, and specify the information source (InfoSource) in the case of a statement. In addition, scene IDs are specified as the object (value of "what") for connections between scenes.

4.8 Mapping verbs to hasPredicate values

- Verb forms are unified in the active voice.
  To facilitate the inference process, the scene "Mr. A was shot by Mr. B" is rephrased as "Mr. B shot Mr. A."
- Verb tenses are unified in the present tense.
  Since time information can be determined by the aforementioned time addition, the verb (the value of hasPredicate) in the scene "Mr. B shot Mr. A" is "shoot" in the present tense, not the past tense.
- Emotional expressions are unified into states, not verbs.
  For example, in the scene "John Straker was excited," "excited" is not treated as a verb, but is taken as a state and the value of hasProperty.
- Scenes involving verbs followed by infinitives are broken down.
  For example, in the scene "John Straker tried to go check the stable", instead of creating a verb like tryTo, we break the scene down as follows.

```
# John Straker tried to go to check the stable.
kd:id-a kgc:subject: kd:John_Straker ;
        kgc:hasPredicate: kdp:try ;
        kgc:what kd:id-x .
# John Straker go to check the stable.
kd:id-x: kgc:subject kd:John_Straker ;
        kgc:hasPredicate kdp:go ;
        kgc:where: kd:the_stable ;
        kgc:why kdp:check .
```

- Auxiliary verbs and verbs concatenated into one verb.
  For example, in the scene "Percy Trivellian had to prepare the money," mustPrepare is created as a verb, and the inference process is facilitated by separately defining that it consists of "must" and "prepare".

4.9 Unification of words such as object and complement

- Assign unique names and IRIs to people and things.
  List the people and things that appear first, and assign unique names and IRIs to them.
- Replace collation with proper nouns and scene IDs.
  To distinguish whether it is a concrete person or thing, replace directives, pronouns, and so forth, with proper nouns.
- Unified notation for labels and IRIs.
  Establish conventions for the use of camel notation, snake notation, space delimiters, and so forth, to ensure consistency within a knowledge graph.

4.10 Uniform treatment of modifiers

Since a modifier may be used as a keyword in a story, we use a resource as it is if it has a qualifier, such as "red carpet." The type is then defined as "carpet" and the property (value of hasProperty) is defined as "red".

4.11 Verification of the guideline application

To apply this guideline, we first conducted a trial application with a third party using one of the target stories, "The Resident Patient", as an example, and examined the costs and procedures required for the work. Specifically, a software engineer who had knowledge of RDF and an outline of the inference challenge, but was not involved in the creation of the guidelines (hereafter, the worker), undertook steps 4.1—4.9 while referring to the guidelines, and summarized considerations for a full-fledged application. The tenth guideline was excluded from the application because it requires the development of a vocabulary to be used as modifiers. Afterwards, the results of the work were shared and discussed among the workers and the guideline authors, and the following findings were obtained.

- The work taken to execute steps 4.1, 4.2, and 4.4 is very costly because each of these guidelines requires close examination based on an understanding of the short sentences that describe the scene and the content of the original novel. For this reason, we decided to extract the scenes to which these guidelines should be applied first.
- For step 4.3, this can be handled by extracting scenes with time entries, finding a reference date and time entry, and shifting the time by several hours from that point as the story progresses.
- For steps 4.5, 4.6, 4.8, and 4.9, it seems possible to extract the target locations and handle them to some extent by automatic processing.
- For step 4.7, since it is costly to change the description of nested expressions, scene types (Table 1) should be given thoroughly and the need for nested expressions should be clearly stated.

The same worker then revised the remaining seven stories based on the guidelines, in accordance with the above policy. RDF triples were converted to a spreadsheet format as the working data, and a comparison tool was developed and used to unify the vocabulary. The approximate time for knowledge graph revision of all stories was about 30 days per person. Finally, the guideline developers reviewed and revised the results. Through this verification, we were able to confirm that the application of the guidelines by a third party was generally appropriate, although some of the work by the third party required modification. It was also found that the steps that were related to the extraction of correction points because of the high time costs included items that should be individually detailed in the correction policy; thus, further studies are needed. Regarding the applicability of this guideline to knowledge graphs in general, we found that steps 4.2-–5 and 4.7 are specific to scene graphs or event-centric knowledge graphs, while steps 4.1, 4.6, and 4.8–10 are common to general knowledge graph refinement.

The constructed and refined knowledge graph is available as open data on the project website[3].

5 KNOWLEDGE GRAPH REASONING CHALLENGE

As mentioned above, this task is to correctly identify the culprit and causes of incidents and accidents using inference and estimation techniques. However, since it can be generalized as a kind of knowledge graph completion[4], it can be positioned as a generic problem setting that can be applied to the construction of various knowledge bases including knowledge graphs, information extraction and relation extraction, knowledge updating and maintenance, and so on. Moreover, in addition to the focus on real social problems and the emphasis on explainability of the results, there are some unique difficulties, such as described in the following points:

- Real-world problems are all individual cases, and similar scenes do not necessarily appear more than once. Therefore, knowledge or data is not necessarily big data, making learning difficult.
- Rather than explaining single relationships by approximation in vector space, they must be assembled or chained together to derive the goal as a whole.
- The knowledge graph includes false statements spoken by the characters.

5.1 Outline of proposed techniques

The following three categories were selected for application:

(1) Main track: develop a system to solve one or more tasks of the target stories.
(2) Tool track: develop tools to solve partially any of the tasks (e.g., suspect estimation, alibi verification, motive explanation, and so forth).
(3) Idea track: derive ideas on how to realize any of the above (possibly without system implementation).

The total number of proposals from the 1st to the 3rd contests was 24 (11 in the main track, 5 in the tools track, and 8 in the ideas track)[5]. This section discusses the overall trend of these proposals. The approaches to the challenge can be broadly divided into the following two categories:

- Knowledge processing approaches to reasoning based on rules using first-order predicate logic, ontology definitions, and so forth.
- Machine learning approaches to finding information that leads to identifying the culprit by learning from the provided knowledge graph, other cases, and novels as training data.

Therefore, the first perspective for comparing the proposed techniques is: (1) whether the method is centered on knowledge processing/machine learning or both. The knowledge graphs of mystery stories provided in this challenge were created by extracting the main parts from the contents of the stories and following the descriptions of the stories. Therefore, the knowledge graphs include a large amount of knowledge that is common knowledge to the readers of the stories and not explicitly described (e.g., "death by a knife to the heart"). For this reason, the challenge allows applicants to supplement external knowledge necessary for inference, and the introduction of useful external knowledge is also an important aspect of evaluation. Therefore, as the second perspective, (2) we compared the external knowledge used in each method. Table 3 shows the results of the comparison of the proposed techniques, focusing on the above two perspectives.

Regarding the type of approach in (1), most of the works in the first challenge focused on knowledge processing. However, we saw a significant increase in the number of approaches using machine learning from the second challenge. In addition, methods that use both approaches in a mutually complementary manner (the main track #1 and #4 in the second challenge) could also be seen, and this suggests that the tasks in this challenge could be a research target that encourages the integration of knowledge processing and machine learning.

On the other hand, for (2), many of the methods in the first challenge used knowledge necessary for inference as originally created ontologies or rule descriptions. However, after the second challenge, many methods that utilize existing resources such as WordNet, Wikipedia, and Wikidata were proposed. It is thought that there is a desire to reduce the cost of describing proprietary knowledge. In the future, it will be an important issue to what extent existing resources can be used as external knowledge.

The details of individual proposals and their evaluation were discussed in[5, 6, 8, 10, 14, 15].

6 SUMMARY AND FUTURE ISSUES

In this paper, we described our findings in constructing and refining scene knowledge graphs using a test set with the goal of contributing to the technological advancement of XAI. Although other knowledge graphs with more triples exist, many of them contain only simple relationships. The knowledge graphs in this study are characterized by including more complex relationships that reflect the real world, such as temporal, causal, and probabilistic relationships. Although there are many studies that automatically generated knowledge graphs from text data using natural language processing techniques, much of the data cannot be used as is. There has been no other study that summarized refinement methods for actually using the constructed knowledge graphs for inference and machine learning. We hope that Section 4 will serve as a general guideline for other studies.

The first issue to be addressed is the organization of the "is-a" hierarchy of verbs. Since this is expected to be a huge task that is also connected to the development of the Japanese language system, we first developed a lexicon for verbs, and then developed a tool to group words with similar semantics. However, there is a tradeoff in combining subtle differences in semantics into a common vocabulary. While such an approach may facilitate symbolic inference, it is known that the details of semantics expressed using embedding and other methods may be lost in some cases. This is one topic that should be considered in the integration of logical reasoning and machine learning. The second is the consideration

[3]https://github.com/KnowledgeGraphJapan/KGRC-RDF
[4]e.g., to regard a particular person as lacking the attribute of a culprit.
[5]The fourth contest was a challenge for students only, so it is omitted here.

Table 3: Comparison of proposed techniques from the knowledge graph reasoning challenges

#	Proposal	Affiliation of applicant	Description	(1) Approach	(2) Ex. knowledge
1	Main 1	Nomura Research Institute, Ltd.	Search for situations that satisfy the content of the testimony based on tensor decomposition and SAT problems	Knowledge	original rules, etc.
	Main 2	Fujitsu Limited	Reasoning using ontologies that describe knowledge about motives, opportunities and means, and inference rules (SHACL)	Knowledge	original ontologies and inference rules
	Main 3	Fujitsu Limited	Estimating the culprit based on machine learning from other stories.	ML	other story text by Sherlock Holmes
	Main 4	Ritsumeikan University	Inference of culprit characters using original rules and ontology	Knowledge	original rules (presumption of culpability) and ontology (motive)
	Main 5	The University of Electro-Communications	The necessary rules are written in a form that can be deduced by the triple store (Stardog) and used to infer the culprit.	Knowledge	original rules, etc.
	Idea 1	Nagoya Institute of Technology	Proposal of the idea of deducing who is the culprit by discussing it with multiple agents, with explanatory	(neither)	none
	Idea 2	Suuri-sentangijutsu Laboratory	Consideration of items necessary to deduce and explain the culprit	Knowledge	original knowledge
	Idea 3	FUJIFILM Business Innovation Corp.	Describe the events and circumstances related to the case based on a reasoning ontology of your own creation, and generate inferences and evidence for the crime	Knowledge	original ontologies
2	Main 1	Nomura Research Institute, Ltd.	Combination of fact extraction by judging similar sentences of criminal frames and BERT, tentative inference using predicate logic, etc.	Knowledge + ML	Text of the stories, original rules, external information
	Main 2	KDDI Research, Inc.	Inference of culpability using knowledge graph embedding techniques	ML	none
	Main 3	Sakiyomi AI Lab Ltd.	Embedding the hyperbolic space of the knowledge graph and visualizing the inference process using the ALBERT model	ML	none
	Main 4	Fujitsu Limited	Extend the previously-created ontology (motive, means) using generic knowledge (WordNet and Wikipedia), and learn and predict knowledge graph embedding	Knowledge + ML	original ontology, WordNet, Wikipedia
	Tool 1	Tokyo City University	Estimation as an unknown entity problem using graph neural networks	ML	none
	Tool 2	Hosei University	Extracting words linked to the culprit using tfidf's important words extraction tool	(neither)	none
	Tool 3	Osaka Electro-Communication University	Mapping the knowledge graph to a dictionary of word emotion intensities to infer changes in characters' emotions from scene to scene	Knowledge	NRC Emotion / Affect Intensity Lexicon
	Idea 1	Fujitsu Limited, Kobe Tokiwa University, Kobe City Nishi-Kobe Medical Center	Examination of "Dancing Men" in light of mathematical methods for deciphering the code.	ML	Wikipedia
	Idea 2	(Individual)	Investigation of a method for guessing the culprit using natural language processing	ML	none
3	Main 1	Tokyo City University	Learning by Continuous Bag-of-Words (CBoW) of word co-occurrence and Principal Component Regression Analysis	ML	none
	Main 2	KDDI Research, Inc.	Predicting missing links in the knowledge graph to estimate the culprit	ML	ConceptNet
	Tool 1	(Individual)	Focusing on time information, the absolute time of an event and the relationship before and after the event are acquired, inferred, and visualized in chronological order	Knowledge	none
	Tool 2	Fujitsu Limited	Simple construction of a novel ontology using Wikidata's class hierarchy	Knowledge	Wikidata
	Idea 1	Panasonic Corporation	Automatic identification of polysemy using similarity with verbs in WordNet (using word2vec and BERT)	ML	WordNet
	Idea 2	Osaka Electro-Communication University	Reasoning with Wikidata and ICD-10 mapping information	Knowledge	Wikidata, ICD-10
	Idea 3	Sakiyomi AI Lab Ltd.	Estimation of knowledge graphs using embedding into the hyperbolic space of knowledge graphs to account for the habits of the creator of the knowledge graphs	ML	none

of new scene representation methods. In the current schema, each scene is represented as a knowledge graph. However, it is also possible to prepare a knowledge graph that represents the situation at each point in the story, and to represent each scene as a differential change to the knowledge graph. This is the difference between Motion JPEG and MPEG. We will consider this is as a future issue.

The topic of XAI using knowledge graphs has become a research topic in recent years, with several workshops and special sessions being held at prominent conferences. In this study, we hosted the first International Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence (KGR4XAI)[7][6] in conjunction with IJCKG2021 in FY2021, and attracted more than 80 participants. In FY2022, we are planning to hold the first International Knowledge Graph Reasoning Challenge. We look forward to the participation of all related parties.

[6]https://kgr4xai.ikgrc.org/

ACKNOWLEDGMENTS

We would like to express our gratitude to all the participants in the workshops, technical meetings, and other events that have been held so far. This work was supported by JSPS KAKENHI Grant Number 19H04168.

REFERENCES

[1] Shusaku Egami, Satoshi Nishimura, and Ken Fukuda. 2021. Virtual-Home2KG: Constructing and Augmenting Knowledge Graphs of Daily Activities Using Virtual Space. *20th International Semantic Web Conference (ISWC 2021), Posters, Demos and Industry Tracks* (2021).

[2] Johannes Frey, Kay Mueller, Sebastian Hellmann, Erhard Rahm, and Maria-Esther Vidal. 2019. Evaluation of metadata representations in RDF stores. *Semantic Web* 10, 2 (2019), 205–229. https://doi.org/10.3233/SW-180307

[3] Simon Gottschalk and Elena Demidova. 2019. EventKG - the Hub of Event Knowledge on the Web - and Biographical Timeline Generation. *Semantic Web J.* 10, 6 (2019), 1039–1070.

[4] Saiping Guan, Xueqi Cheng, Long Bai, Fujun Zhang, Zixuan Li, Yutao Zeng, Xiaolong Jin, and Jiafeng Guo. 2022. What is Event Knowledge Graph: A Survey. *IEEE Transactions on Knowledge and Data Engineering* (2022). https://doi.org/10.1109/TKDE.2022.3180362

[5] Shuhei Katsushima, Hajime Anada, Shusaku Egami, and Ken Fukuda. 2021. A Criminal Detection of Mystery Novel Using the Principal Components Regression Analysis Considering Co-Occurrence Words. *1st Int'l Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence (KGR4XAI)* (2021).

[6] Takahiro Kawamura, Shusaku Egami, Koutarou Tamura, Yasunori Hokazono, Takanori Ugai, Yusuke Koyanagi, Fumihito Nishino, Seiji Okajima, Katsuhiko Murakami, Kunihiko Takamatsu, Aoi Sugiura, Shun Shiramatsu, Xiangyu Zhang, and Kouji Kozaki. 2019. Report on the First Knowledge Graph Reasoning Challenge 2018 - Toward the eXplainable AI System -. *9th Joint Int'l Semantic Tech. Conf. (JIST2019)* (2019), 18–34.

[7] Takahiro Kawamura, Boris Villazón-Terrazas, Marut Buranarach, and Kouji Kozaki. 2021. Summary of the 1st International Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence. *1st Int'l Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence (KGR4XAI)* (2021).

[8] Kouji Kozaki, Shusaku Egami, Kyoumoto Matsushita, Takanori Ugai, and Takahiro Kawamura. 2021. Knowledge Graph Reasoning Techniques through Studies on Mystery Stories -Report on the Knowledge Graph Reasoning Challenge 2018 to 2020. *1st Int'l Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence (KGR4XAI)* (2021).

[9] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. 2017. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International journal of computer vision* 123, 1 (2017), 32–73.

[10] Mori Kurokawa. 2021. Explainable Knowledge Reasoning Framework Using Multiple Knowledge Graph Embedding. *10th International Joint Conference on Knowledge Graphs* (2021).

[11] Vincenzo Lombardo, Rossana Damiano, and Antonio Pizzo. 2018. Drammar: A Comprehensive Ontological Resource on Drama. *17th Int'l Semantic Web Conf. (ISWC2018)* (2018), 103–118. https://doi.org/10.1007/978-3-030-00668-6_7

[12] Maroua Masmoudi, Sana Ben Abdallah BenLamine, Hajer Baazaoui Zghal, Bernard Archimede, and Mohamed Hedi Karray. 2021. Knowledge hypergraph-based ap- proach for data integration and querying: Application to Earth Observation. *Future Generation Computer Systems* 115 (2021), 720–740.

[13] Marco Rospocher, Marieke van Erp, Piek Vossen, Antske Fokkens, Itziar Aldabe, German Rigau, Aitor Soroa, Thomas Ploeger, and Tessel Bogaard. 2016. Building event-centric knowledge graphs from news. *J. of Web Semantics* 37-38 (2016), 132–151. https://doi.org/10.1016/j.websem.2015.12.004

[14] Takanori Ugai. 2021. Fuzzy Search of Knowledge Graph with Link Prediction. *10th International Joint Conference on Knowledge Graphs* (2021).

[15] Takanori Ugai, Yusuke Koyanagi, and Fumihito Nishino. 2021. A Logical Approach to Criminal Case Investigation. *1st Int'l Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence (KGR4XAI)* (2021).

[16] Willem van Hage and Davide Ceolin. 2012. The Simple Event Model. *Situation Awareness with Systems of Systems* (2012), 149–169. https://doi.org/10.1007/978-1-4614-6230-9_10

[17] Paul Vicol, Makarand Tapaswi, Lluís Castrejón, and Sanja Fidler. 2018. Moviegraphs: Towards understanding human-centric situations from videos. *IEEE Conference on Computer Vision and Pattern Recognition* (2018), 8581–8590.