International Knowledge Graph Reasoning Challenge (IKGRC)

Application Sheet

*Submit a PDF file containing the contents of this application sheet by e-mail.
(Any software can be used to create the sheet)

Contact: kgrc@knowledge-graph.jp

- Applicants to Main and Tool Tracks must describe all the information from 1 to 3.
- Applicants to Idea Track must describe at least information 1 and 2.
- There is no limit to the number of pages, so please add more pages as needed.

1. Information about the applicants

- Authors [Name or Team's name]: Kazuma Hasegawa and Akihiro Fujii
- Affiliations: Hosei University
- E-mail (contact): kazuma.hasegawa.5w@stu.hosei.ac.jp

2. Explanation of the reasoning and estimation process

- Explain key points using the results, e.g., program logs. The format is free in the case of Idea Track.
(Please explain clearly with diagrams, etc.)

We developed a tool to support the construction of knowledge graphs for mystery novels using BERT. The purpose of this tool is to automate the process of extracting the parts of a novel to be made into a knowledge graph, which is a part of the procedure of constructing a knowledge graph. We assume that the parts to be made into a knowledge graph are the parts that have important meanings in the novel. The parts that have important meanings in the novel are defined as scenes that represent characters' actions, feelings, and personalities, or incidents and situations in the novel. A summary sentence used in this study is defined as a sentence that contains these important meaningful parts. BERT is a natural language processing model that employs a transformer that can return context-sensitive word and vector representations. This model is used to perform a binary classification of whether the

input sentence is a summary or not (non-summary). Figure 1 shows an overall diagram of the model.

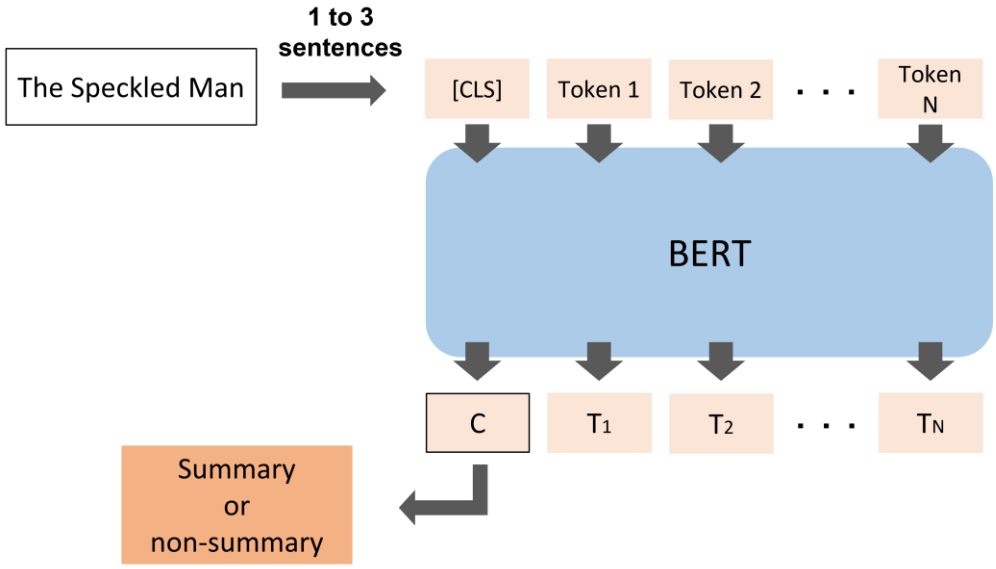

Figure 1 Overall view of the model

As shown in Figure 1, the test will be conducted on the work "The Speckled Band". The training data for the model will be "The Dancing Men," "Silver Blaze," "A Case Of Identity," "The Resident Patient" and "The Crooked Men". The dataset is divided into 1 to 3 sentences each so as not to be too short, which are then input into BERT; the sentence vectors obtained from BERT are used to classify whether the sentence is summary or non-summary. Figure 2 shows the results of training on one work (The Dancing Men) and the results of training on five works.

	Accuracy	Precision	Recall	F-score
One work (The Dancing Men)	0.71	0.61	0.57	0.59

	Accuracy	Precision	Recall	F-score
Five works	0.67	0.58	0.80	0.67

Figure 2 The results of training on one work (The Dancing Men) and the results of training on five works.

The increase in the number of works from one to five allowed BERT to learn a wide variety of expressions. This increased the number of times it determined that an input sentence was a summary, regardless of whether the correct answer was a summary or not. As a result, the reproduction rate increased significantly and the fit rate decreased slightly. This result suggests that training on more works will help BERT understand more about summary sentences and improve its accuracy.

- Range of knowledge graphs used (Scene ID)
The Speckled Band, The Dancing Men, Silver Blaze, A Case Of Identity, The Resident Patient, The Crooked Men

3. Developed application

4. Sharing of materials

If possible, we would like to publish your “Application Sheet” and “Program” on the IKGRC website*. Please let us know by filling out the form below.
*We may use some of the submitted materials in events related to the IKGRC.

[Application Sheet]
- Public availability:
    (○) Allow to publish
    (　) Do not allow to publish
- Publication format
    (○) Published on the IKGRC website
    (　) Publish on your own site and wish to be linked to our site.
      -> URL*:

[Submitted application, source code, data, etc.]
- Public availability:
    (　) Allow to publish
    (○) Do not allow to publish
- Publication format
    (　) Published on the IKGRC website
    (　) Publish on your own site and wish to be linked to our site.

-> URL*:

* The authors may contact us with the URL of the publication website after the website has been opened.