Extraction method of access-related information from reviews of tourist attractions

YUKI SHIODA[1], YOKO NISHIHARA[2]

[1]GRADUATE SCHOOL OF INFORMATION SCIENCE AND ENGINEERING, RITSUMEIKAN UNIVERSITY

[2]COLLEGE OF INFORMATION SCIENCE AND ENGINEERING, RITSUMEIKAN UNIVERSITY

Research Background

Official access information

- Only provides basic information
  (e.g., required time from the nearest station or bus stop)
- Does not describe detailed walking experiences

Actual walking route to Kiyomizu-dera

- Uphill slope from the bus stop
- Many souvenir shops and cafés along the way

Such access-related information is often written only in user reviews on travel websites.

It needs a method to automatically extract and summarize access-related information from reviews.

Our Approach

Example review of Kiyomizu-dera Temple:

Starting keyword (the first word of access related sentences)

From the **bus stop**, the uphill walk to Kiyomizu-dera is quite long, but there are many different shops along the way, so you can enjoy walking without getting bored. When you **arrive**, the view from the stage is so expansive…

Ending keyword (the last word)

We extract the sentences between the starting and ending keywords.

- **Some extracted text is not truly access-related or just mentions a location**(e.g., “Walk to Kiyomizu-dera Temple”)
  - We extract **dependency pairs** (noun–verb, noun–adjective).
  - We keep only partial reviews that contain at least one such pair.
- **We focus on evaluative and experiential expressions about access.**

Research Objective

Objective

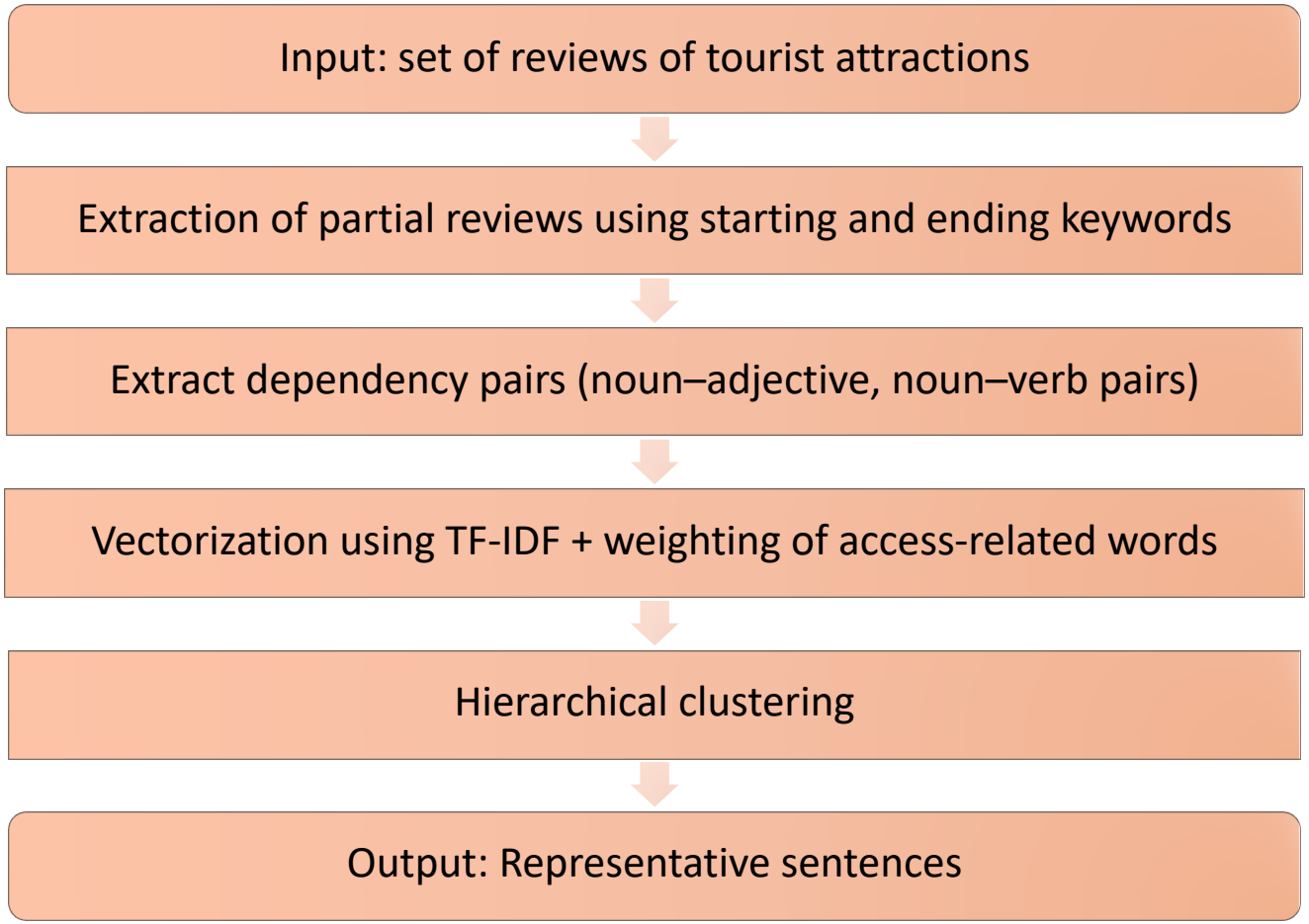
- To extract sentences containing **access-related information** from travel reviews

Method

- The method uses starting and ending keywords to extract partial reviews.
- It removes unnecessary sentences using dependency pair extraction.
- It clusters similar partial reviews and extracts representative sentences.

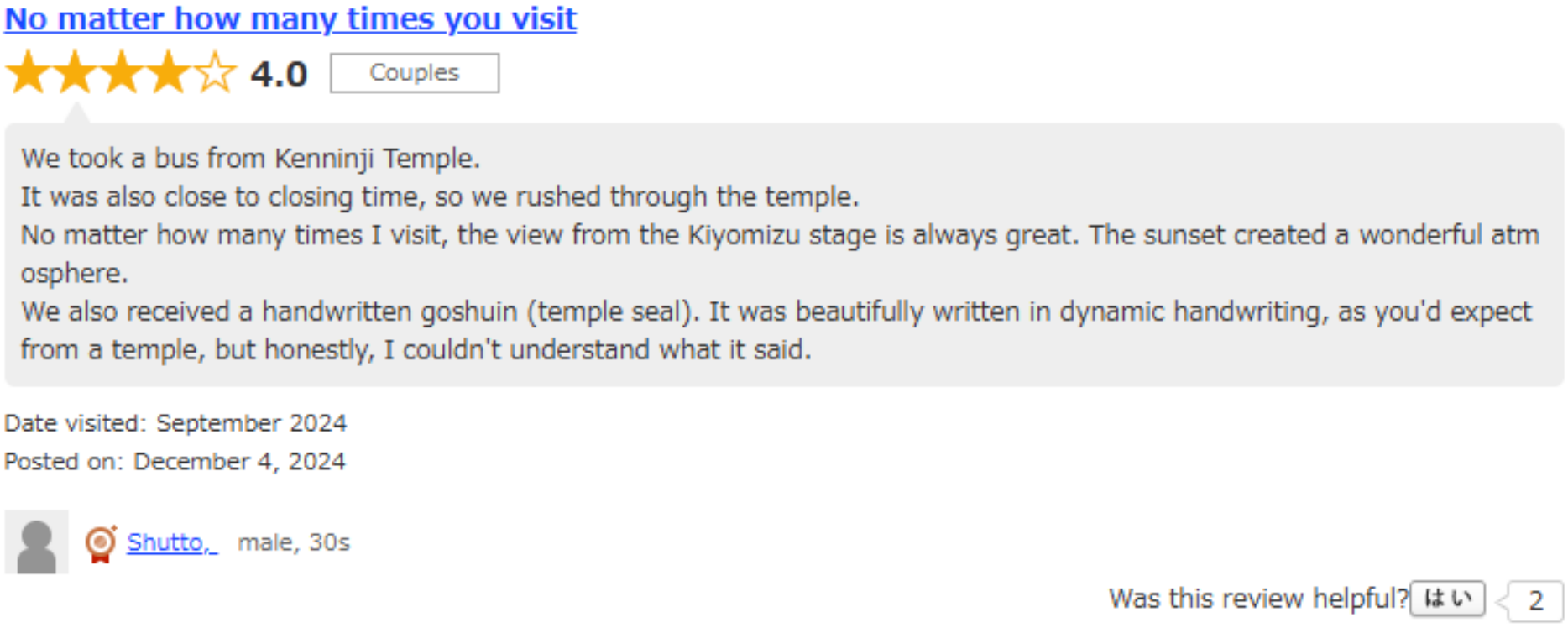
We build a system that summarizes access-related information that is not provided on official websites.

Outline of the Proposed Method



Input: Set of Reviews of Tourist Attractions

The method extracts only the section containing user reviews
from the “**Reviews**” section of the travel site
“**Jalan.net**”(the area with the gray background in the image below).



Extraction of Access-Related Review

Based on manually extracted access-related reviews,
the method determines the starting keyword and ending keyword.

Starting keywords

- “station,” “bus stop,” or mode of transportation.

Ending keywords

- Tourist attraction names, specific locations within attractions, arrival notifications.

Starting Keywords	Kiyomizu-michi, Kyoto Station, Kiyomizu-Gojo Station, from the station, bus stop, Gion-Shijo Station, nearest bus stop, on foot, by walking, by taxi, by car, by bus, by train, by bicycle, to Kiyomizu-dera, to get to, to Kiyomizu-dera, approach road, along the way
Ending Keywords	Kiyomizu-dera, arrival, arrived, will arrive, have arrived, to arrive, stage, precincts, can go, went, managed to reach, got there, get to, recommended

Extraction Dependency Pairs

Purpose

- To capture **evaluations and experiences** about access
- To remove partial reviews that contain only non-informative fragments

Extracted pairs

Noun–verb and noun–adjective combinations

Examples:

- slope - steep
- souvenir shops – enjoy

- **We focus on evaluative and experiential expressions about access**

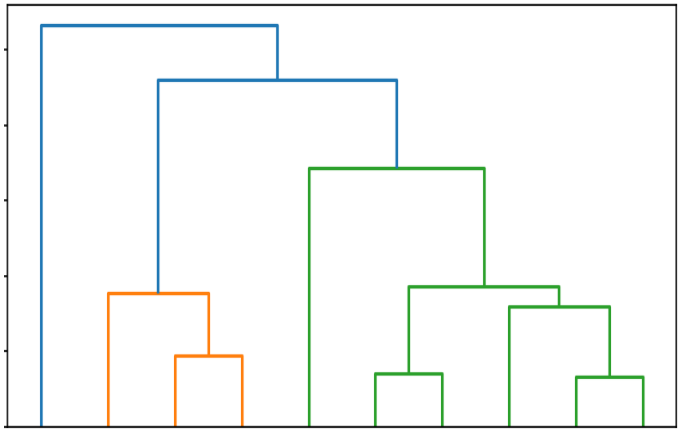
Hierarchical Clustering

Purpose

Similar reviews are grouped together to reduce the number of reviews presented to the user.

Method

- Reviews are vectorized using **TF–IDF.**
- Higher weights (× 3) for **access-related words** (e.g., “bus stop”, “on foot”, “slope”).
- Hierarchical clustering (Ward’s method)
- Distance threshold controls the number of clusters.



Hierarchical clustering

Evaluation Experiment

Objective

- To evaluate the effectiveness of the method

Procedure

- The method performs hierarchical clustering (Ward's method).
- We set the distance threshold 2.2, which can obtain better clustering results.
- We manually judge whether each cluster represents useful access-related information.
- For each tourist attraction, we compute the number of clusters and the precision.

Data

Tourist attraction	Number of reviews
Kiyomizu-dera Temple	4,840
Fushimi Inari Taisha	4,837
Heian-jingu Shrine	1,066
Kinkaku-ji Temple	2,095
Ginkaku-ji Temple	1,185

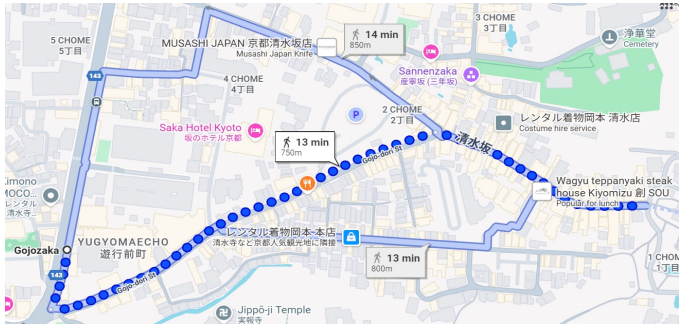
Experimental Results: Precision Evaluation

	Number of clusters	Precision
Kiyomizu	81	0.69
Fushimi Inari	17	0.76
Heian Jingu	4	0.75
Kinkaku	3	0.66
Ginkaku	6	0.50
Average	22.2	0.672

We obtained 0.672 precision on average.

- For tourist attractions with diverse routes (e.g., Kiyomizu-dera Temple)

→ Many reviews are obtained, and the method is effective.



- For attractions with simple routes (e.g., Kinkaku-ji Temple)

→ Access-related reviews are less, which leads to lower precision.

Experimental Results: Extracted Review Examples

Correctly extracted reviews:

Describing **walking experiences**:

- presence / absence of slopes or stairs
- shops and cafés along the route
- time required from bus / subway stations

Incorrectly extracted reviews:

Not truly about **walking access**:

- impressions of atmosphere
- ticket or pass information
- access inside large attractions

Tourist attraction	Correctly extracted reviews
Kiyomizu-dera	There are **many souvenir shops** along the way.
Heian-jingu Shrine	The **bus stop is close**, and it's about a 10-minute walk from the subway, so access is good.
Kinkaku-ji Temple	The walk there is also **quite impressive and worth seeing**.

Tourist attraction	Incorrectly extracted reviews
Heian-jingu Shrine	There are terrace seats, so when you get tired of sightseeing, **you can relax with a coffee** while viewing Heian-jingu.
inkaku-ji Temple	You can go by bus, and I also recommend the **two-day subway and bus pass**.

Conclusion

Summary

- We proposed a method to extract **access-related information** from reviews.
- The proposed method extracts partial reviews using **starting and ending keywords.**
- The method applies dependency pair filtering and hierarchical clustering.

Results

- We evaluated the method on **14,023 reviews** from **five tourist attractions.**
- We achieved **better precisions** in extracting access-related information.

- The proposed method can **effectively extract access-related information** from user reviews.