Design and Evaluating a Method Using Project Corpus for Inferring Software Description

1st Kohei Terakawa
Kobe University,
1-1 Rokkodai-cho, Nada,
Kobe, 657-8501, Japan
odajin@ws.cs.kobe-u.ac.jp

2nd Sinan Chen
Kobe University,
1-1 Rokkodai-cho, Nada,
Kobe, 657-8501, Japan
chensinan@gold.kobe-u.ac.jp

3rd Masahide Nakamura [1,2]
[1]*Kobe University,*
[2]*RIKEN Center for Advanced Intelligence Project,*
Tokyo, 103-0027, Japan
masa-n@cs.kobe-u.ac.jp

***Abstract*—Obsolete software developed in the past is gradually phased out over time. However, the source code of such software contains a wealth of information that can be re-purposed and possesses a high value as an asset. Thus, understanding the characteristics of existing software can aid in developing new software. In a previous study, we proposed a method for inferring the architecture of an existing system using a project corpus and conducted preliminary experiments to verify its feasibility. The findings revealed that the project corpus could be used to infer a system's purpose, functionality, and technology. In this present study, we confirm the validity of the project corpus from a perspective that was not examined in previous studies. We established three verification items and conducted an experiment in which we employed project corpus to infer the functionality of the system, the technology utilized in the system, and the architecture of the system. From the experiment results, we confirmed that the accuracy of the project's inference depends on two factors: first, that the project corpus accurately reflects the system's information, and second, the participants' familiarity with the words in the corpus.**

***Index Terms*—software development, technical debt, upcycling, corpus**

I. Introduction

The gradual obsolescence of software developed in the past can be attributed to various factors, such as advances in the development environment and the accrual of technical debt [1]. However, the source code of such software contains a wealth of reusable and valuable information that can be leveraged in developing new software by comprehending the characteristics of the existing software. In order to recycle existing software products, we are exploring the concept of software upcycling. In software upcycling, it is crucial to comprehend a pre-existing project's objective and functionality and discover reusable concepts and designs.

In a previous study, we proposed a method of inferring the purpose and function of a system using a project corpus, $corpus_{Pj}$, thereby providing an understanding of the outline of an existing project. Preliminary experiments were conducted to verify the effectiveness of the proposed method. However, **certain challenges have persisted in the initial experiments** which were carried out by three participants. However, it is imperative to augment **the number of participants** and conduct additional experiments to investigate the efficacy of the proposed technique. Additionally, there is scope for refinement of the experimental methodology. Initial experiments indicated that technical insights could be gleaned from $corpus_{Pj}$. Therefore, it is imperative to validate the effectiveness of the project corpus from the perspective of comprehending the technology employed.

This study aimed to ascertain the validity of $corpus_{Pj}$ from an alternate perspective than that of prior research. To achieve this goal, we formulated the following research question: (RQ1) Is it possible for $corpus_{Pj}$ to discern the system's functionality? (RQ2) Is it feasible for $corpus_{Pj}$ to provide an understanding of the technology employed in the system? (RQ3) Could $corpus_{Pj}$ provide a comprehensive understanding of the system? Moreover, the survey will be administered to more participants than in preceding studies, and the findings will be analyzed.

In our experiment, the $corpus_{Pj}$ employed is constructed utilizing the method proposed in the preceding study. $corpus_{Pj}$ is a denomination stipulated in the proposed methodology to denote a corpus within the context of software. A corpus [2] is typically utilized in linguistics and refers to an extensive database of texts and utterances. In other words, a corpus in software is a summary of a project created from the information contained in the project. The proposed method of the previous study concentrated on class name component words as key concepts. Class names are assumed to be closely associated with the objective and functionality of a project, and class name component words are extracted. The resulting words are deemed to embody the essence of the project from. They are extracted and are used as a reference to infer the outline of the project. The experiment comprises the following five steps: (Step 1) Selection of project and participants. (Step 2) Corpus preparation. (Step 3) Questionnaire design. (Step 4) Participants' responses. (Step 5) Evaluation of responses.

The experiment's findings indicated that there were queries with a substantial proportion of accurate answers and queries with a minimal proportion of accurate answers. The analysis results affirmed that two principal factors influenced the comprehension of the project outline: (1) $corpus_{Pj}$ must contain adequate information about the system. (2) It transpires that a certain degree of familiarity with pre-existing technology and some background in software engineering is necessary to make deductions.

979-8-3503-0050-5/23/$31.00 ©2023 IEEE

II. Previous Study: Reasoning Existing Projects' Descriptions Based on Classname Word Elements [3]

A. Research Focus

Software upcycling [4] is a technique of software reuse. Software upcycling transforms existing software products into valuable software assets (code, services, etc.) by focusing on their design and implementation and incorporating novel modifications. Consequently, in software upcycling, it is crucial to comprehend the objective and functionality of a pre-existing project and discover reusable concepts and designs [5]. However, not all software products in organizations and companies are maintained in an asset-like fashion. In certain instances, documentation such as $README$ is absent or inadequate. In order to foster upcycling, it is necessary to have an overview of the project regardless of the existence or otherwise of documentation.

B. Proposed Method

In the previous study, we proposed a methodology for obtaining a comprehensive perspective of a pre-existing project without relying on the $README$ or other documentation. The $corpus_{Pj}$ is constructed from class name constituent words, and the objective and functions of the system are deduced from it. The outline is presented in Fig. 1. It exemplifies applying the proposed methodology to a project titled MP3PlayService. In step 1, repository mining is executed to acquire all class file names of extant classes, and in step 2, the class names obtained are partitioned into words, and the resulting word group is denoted $corpus_{Pj}$. In step 3, the words obtained in Step 2 are arranged in lexicographic order, and a tag cloud [6] image is generated utilizing weights based on word frequencies. In Step 4, the objective and function of the original project are deduced utilizing $corpus_{Pj}$ expressed in two distinct manners.

C. Project Corpus

The proposed method defines $corpus_{Pj}$ as a set of words obtained through the class name segmentation process. A corpus, commonly used in linguistics, refers to a collection of linguistic data, typically comprising a database of sentences drawn from various sources such as books, articles, and other written materials. That is, $corpus_{Pj}$ denotes a corpus in the context of software. Specifically, class names incorporated in a project are amassed and separated into words. The acquired words are deemed to reflect the character of the project from which they are extracted, and if the project outline can be deduced from these words, it is effortless to obtain a comprehensive perspective of the project, even if the README is absent.

D. Results of Preliminary Experiment

A preliminary experiment utilizing the proposed method was conducted on three participants. All participants correctly answered more than half of the questions, indicating that the class name component words significantly impacted comprehension of the project outline. However, there are limitations to the study. The preliminary experiment was conducted with a small sample of three participants. Thus, it is essential to conduct further experiments with a larger sample size to assess the proposed methodology's effectiveness fully. Discussion of the preliminary experiment revealed that $corpus_{Pj}$ provides technical clues. Some participants provided a mixture of descriptions of functions and technology used, and it was impossible to isolate the degree to which they understood technical matters. Therefore, it is necessary to verify the validity of $corpus_{Pj}$ from the viewpoint of understanding the technology used.

III. Experimental Design

A. Purpose of Experiment

The goal of this study is to verify the effectiveness of the proposed method from a perspective not considered in previous studies. For this purpose, we set the following research questions: **RQ1:** Is it possible for $corpus_{Pj}$ to discern the functionality of the system? **RQ2:** Is it feasible for $corpus_{Pj}$ to provide an understanding of the technology employed in the system? **RQ3:** Is it possible for $corpus_{Pj}$ to provide a comprehensive understanding of the system?

B. Experiment Overview

An experiment is conducted in order to address the research questions. The experiment comprises the following five steps: **(Step 1) Selection of Projects and Participants:** 154 projects managed within our laboratory is selected as participants. **(Step 2) Corpus Creation:** A $corpus_{Pj}$ is created utilizing the method proposed in a previous study. **(Step 3) Question Design:** Three questions is formulated to address the research questions. **(Step 4) Participant Responses:** Participants is asked to provide their answers in natural language using a Google Form. **(Step 5) Scoring of Answers:** Individuals with knowledge and understanding of each project evaluate and score the responses.

C. STEP 1: Selection of Project and Participant

The projects utilized in this experiment are Java-based projects hosted on Gitlab, which our laboratory maintains. A random sampling of ten projects is selected from these projects and employed as experimental data. The specifics of the projects employed in the experiments are presented in Table I, which highlights the details of the projects utilized in the experiments. The summary and functions presented in Table I were created by knowledgeable Gitlab administrators for each project, who extracted the main functions and created a summary. For this experiment, it is desirable to have participants with similar conditions such as knowledge and experience. Therefore, for this study, 12 students majoring in information science at the same university were selected as participants.

D. Step 2: Corpus Creation

The proposed method from the previous study is utilized to construct $corpus_{Pj}$. Repository mining is executed on ten experimental projects to extract class name component words using GitlabAPI. The obtained data is

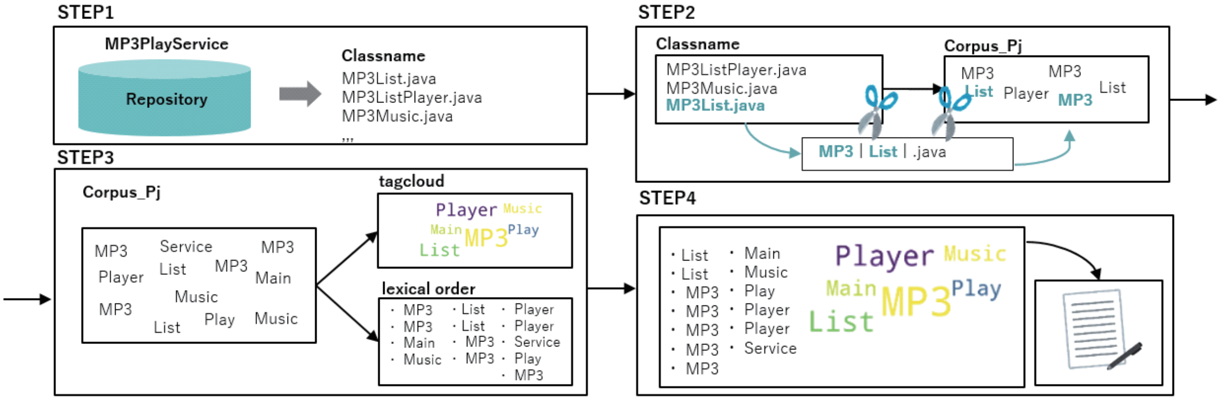


Fig. 1. Flow of the proposed method in the previous study.

TABLE I
DETAILS OF PROJECT UTILIZED IN THE EXPERIMENT.

No.	Summary	Function	Technology used
1	**NewLINEAgentService** service that sends LINE messages using LINE Bot	・Send LINE Messages ・Keep a log of messages	Webhook, LINE Messaging API, Jersey, MongoDB, NextCloud
2	**KobeDemographicAPI** API to obtain population data for each town and street in Kobe City	・Obtain population data ・Get town code	Jersey MySQL
3	**LifeActivityMailer** Email notification service when there is a change in environmental sensors	・Watch for sensor changes ・Notify the owner (administrator) of the sensor by email when a change occurs	JavaMail Sensor box change point detection service(SBCPDService)(SBCPD = SensorBox Change Point Detection)
4	**BLEAdapter** Adapter that formats BLE (Bluetooth Low Energy) data and adapts it to the A-framework	・Analyze raw BLE data, convert to standard IDs, and post to the A-DB .・Post the given location in the A-DB.	Java Web Service XML
5	**enishi_mock** Mock of Enishi mobile application Enishi is a job matching service on the street using a passing-by framework.	・AAndroid Mobile Application Mock ・Screen design only	Android SDK
6	**ENISHI** Enishi's back-end services	・User Management Functions ・Job Posting ・Job seeker information acquisition	Jersey Firebase MySQL
7	**MP3PlayService** MP3 playback service based on a given URL Playback by playlist is also available.	・Playing MP3s ・Playing a list of MP3s ・Stopping playback ・Obtaining the URL of the currently playing song ・Advance to the next track	Apache Axis2 JLayer(MP3 playback library)
8	**virtual-care-giver** Web application that executes care scenarios using virtual agents Calling out, reminding the user of forgotten items, schedule management, YouTube playback, etc.	・Execution of care scenarios ・Talking to and interviewing the patient ・Playback of Youtube videos ・Management of patient information	Java WebSocket REST Web API JavaScript
9	**thin-cas** Lightweight command line application to perform Context Aware Service (CAS) with ECA rules	・Registration of ECA Rules ・Evaluation of ECA rules ・Execution of actions	Pure Java
10	**StateCacheService** Web service to cache application state Holds arbitrary state and information as identifier (key) and Json (value) pairs	・State storage ・Get status ・Get all identifiers ・Acquisition of all states	Jersey

then segmented into class names using Python, creating $corpus_{Pj}$ for each of the ten projects. The acquired $corpus_{Pj}$ are arranged in lexicographic order, and a tag cloud image generated utilizing the word frequencies as weights. Table II illustrates the generated $corpus_{Pj}$ and tag cloud images.

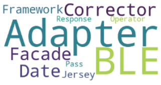
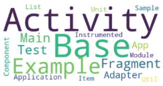
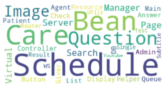
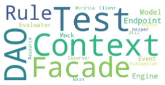
E. Step 3: Design of Questions

The following three questions are designed to assess participants understanding of the project: **(Q1)**: Please list as many bullet points as possible to outline this system's functionality. **(Q2)**: Please outline in bullet points the many technologies and mechanisms you can discern that this system utilizes. **(Q3)**: Summarize in one or two lines a general understanding of the system as inferred from $corpus_{Pj}$. The questions Q1, Q2, and Q3 align with the research questions RQ1, RQ2, and RQ3, respectively. Each question will be examined in Section IV and will serve as a response to the research questions.

F. Step 4: Answers by participants

The $corpus_{Pj}$ generated in Step 2 is presented to participants on a Google Form, and they are asked to input their answer. Two expressions prepared in Step 2 are

TABLE II
CORPUS CREATED IN THE EXPERIMENT.

Question No.	$corpus_{Pj}$	image
1	Beacon Cloud Content Controller DAO DAO DAO DAO Date Drive Event Event Event Event Event Event Event Event Event Follow Handler Handler Handler Handler Handler Handler Handler Handler Http Http Info Info Join LINE LINE LINE LINE LINE LINE LINEBOT LINEBOT LINEBOT Leave Listener Listener Listener Log Log Message Message Message Method Mongo Next Postback Profile Profile Response Service Test Time Unfollow Unknown User User Utils Utils Utils Utils Webhook Webhook Wrapper	
2	1 45 5 Area Connector Demography Demography Error Gson Hello List List Message My Population Rest Rest SQL Test Town Town Town Town bin connector jar mysql	
3	Box Detection Listener Location Mailer Mailer Mailer Mailer Manager Model Notify SBCPD SBCPD Score Sensor Service Service Solver Stream Task	
4	Adapter BLE By Corrector Date Facade Framework Jersey Operator Pass Response	
5	Activity Activity Adapter App Application Base Base Component Example Example Fragment Fragment Instrumented Item List Main Main Module My My Sample Test Test Unit Util	
6	1 1 39 39 5 5 Account Account Account Account Activity Activity Adapter Adapter App Application Application Base Base Client Component Controller Controller Controller Detector Enishi Example Facade Facade Firebase Fragment Fragment Fragment Fragment Fragment Fragment Fragment Info Info Info Info Info Info Info Info Info Info Info Info Info Info Item Jersey List List List List List List List List Main Main Map Matching Messaging Module My My My My Mysql News Offerer Offerer Offerer Offerer Offerer Offerer Parser Passby Passby Passby Sample Seeker Seeker Seeker Seeker Seeker Seeker Service Service Service Test Test Unit User User User User Util View Web Xml bin bin connector connector jar jar mysql mysql	
7	List List MP3 MP3 MP3 MP3 MP3 Main Music Play Player Player Service	
8	Admin Agent Answer Bean Bean Bean Button Care Care Care Check Controller DB Display Giver Helper Http Image Image List Main Manager Manager Page Page Patient Question Question Question Queue Resource Result Router Schedule Schedule Schedule Schedule Search Search Server Server Single Subtitle Test Utils Virtual Virtual WS Youtube	
9	Client Context Context Context Context Context Context DAO DAO DAO DAO DAO DAO Endpoint Endpoint Endpoint Endpoint Engine Engine Engine Evaluation Evaluator Evaluator Event Event Facade Facade Facade Facade Facade Helper Http Http Http Main Mock Mock Model Model Model Morphia Observer Resource Router Rule Rule Rule Rule Rule Rule Test Test Test Test Test Test Test Util	
10	Cache State	

utilized for the $corpus_{Pj}$. To measure the time required and answering, participants were instructed to record their responses' start and finish times.

TABLE III
EVALUATION CRITERIA IN EXPERIMENTS.

Evaluation	Criterion
◎	Correct
○	Mostly Correct
△	Partialy Correct , or Partialy Incorrect
x	Incorrect
?	Unclear

TABLE IV
PERCENTAGE OF CORRECT ANSWERS.

Question Number	Q1	Q2	Q3
1	58%	83%	75%
2	58%	92%	58%
3	92%	67%	67%
4	8%	42%	0%
5	0%	0%	0%
6	42%	83%	25%
7	92%	75%	100%
8	83%	83%	58%
9	58%	42%	42%
10	50%	17%	50%

G. Step 5: Scoring of Answers

The responses obtained in Step 4 are scored by individuals familiar with the details of the project being experimented upon.

One who has a good understanding of the experimental projects conducts the scoring. The evaluation criteria are presented in Table III. The presence or absence of the elements shown in Table I is roughly evaluated based on the following 5-level scale within the answers.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Outline

The experiment yielded responses from all 12 participants. We defined this experiment's percentage of correct answers as the sum of ◎ and ○ scores. The formula is as follows:

$$Correct\,rate(\%) = \frac{Sum\,of\,\circledcirc and \bigcirc}{Total\,number\,of\,answers} \times 100$$

The percentage of correct answers to the three questions Q1, Q2, and Q3 for each question is presented in Table IV. Question 7 had the highest rate of correct answers across Q1 Q3, while Question 5 had a rate of 0 for all questions.

B. (Result 1) Analysis for RQ1

Fig. 2 illustrates the scoring results for each question in response to Q1. Twelve participants provided a rating of ◎, ○, △, ×, or ? Consequently, a total of 12 ratings were obtained for each question. The graph displays the breakdown of the scoring results for each question, with darker areas indicating higher scores. The graph indicates that more than 80% of the participants answered ◎ and

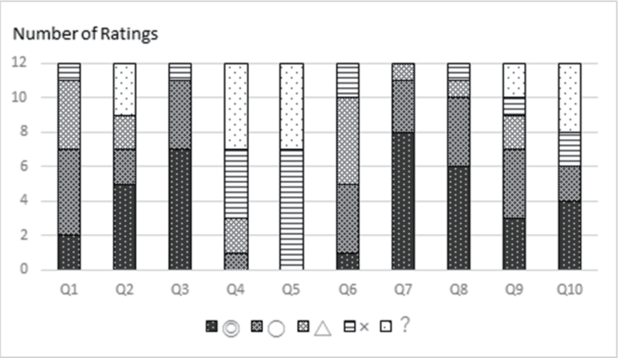


Fig. 2. Breakdown of scoring results for Q1.

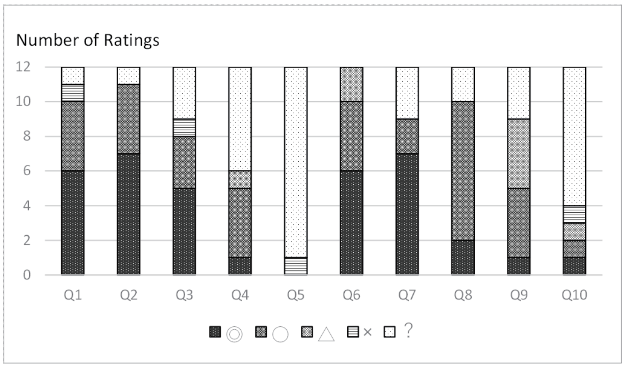


Fig. 3. Breakdown of scoring results for Q2.

○ for questions 3 and 7, indicating a high percentage of accurate responses. Conversely, for questions 4 and 5, the percentage of accurate responses is low, and the scoring results reveal that more than 70% of the answers are marked with X or ? representing a majority of the total number of correct answers. The percentage of accurate responses to questions 4 and 5 could be higher. To examine the cause of the higher percentage of correct answers, we focus on the LifeActivityMailer in question 3. As depicted in the overview, the project comprises two functions: detecting the value of a sensor box and notifying the owner of the sensor via email upon detection of a change. Seven out of twelve participants accurately described these functions, resulting in a score of ◎. The corpus for question 4, as presented in TableII, is listed below.

- Box Detection Listener Location Mailer Mailer Mailer Mailer Manager Model Notify SBCPD SBCPD Score Sensor Service Service Solver Stream Task

The lexicon presented here pertains to the dual functionality of LifeActivityMailer. The lexical items associated with the functions are listed below: (Function 1: Detection of the value of a sensor box) Sensor, Box, Score, Detection (Function 2: Notification of the sensor owner via email) Mailer, Notification

It is intuitive to infer that a sensor detects a value. Thus it is reasonable to assume that most participants focused on the words Sensor, Score, and Detection from $corpus_{Pj}$. Thus, it was revealed that the participants are inferring the system's functionalities by connecting semantically related words together. To investigate the factors contributing to the low rate of correct responses, we focus on question 5. As illustrated in TableI, this project pertains to a screen design for another project. Given that the project was created using the Android SDK [7], the corpus includes component names specific to the Android SDK, such as Activity and Fragment. These concepts are utilized in Android application development and are not germane to the specific functionality of the project. These findings suggest that when the corpus lacks adequate information for inference, it becomes challenging to deduce the functionality.

C. (Result 2) Analysis for RQ2

Fig. 3 illustrates the scoring results for each question in response to Q2. It can be observed that for questions 1, 2, 6, 7, and 8, a majority of answers were scored as ○ and ◎, indicating a high percentage of correct answers. Conversely, questions 4, 5, and 10 exhibited a low percentage of correct answers, with a conspicuous presence of X or ? scores. Most participants provided correct answers by referencing specific terminology within the corpus and utilizing analogies to describe functions. For instance, consider Question 6. As shown in the TableI, the function of Question 6 is to exchange user information with a database. 10 of the 12 participants listed MySQL as the technology they use, and at the same time, they wrote about CRUD in their answers to Q1. The presence of the word "MySQL" suggests that the participants interpreted that the system has a function to perform CRUD to the DB. The above results indicate that when words related to the technology used are present in the corpus, participants use technical terms to infer the system's behavior. On the other hand, some examples were interpreted differently by different participants. In Question 4, more than half of the responses were answered ?(Unclear). As shown in TableI, the function of Question 4 is to post the data obtained by BLE to the DB. BLE [8] is one of the extensions of Bluetooth. It can communicate at meager power and is used for various sensors. However, only three participants described BLE as a technology used and read the function of BLE as a sensor. The amount of information that can be read from the technology used was found to vary from participant to participant.

D. (Result 3) Analysis for RQ3

Fig. 4 illustrates the scoring results for each question in response to Q3. As opposed to the graph of RQ1, questions with high and low percentages of correct answers are prevalent. However, the percentage of correct answers for questions 3, 6, 8, and 9 is lower than that for Q1. The results of our examination of the virtual-care-giver in question 8, which exhibited a lower percentage of correct answers than in Q1, reveal that the system comprises various types of services provided by a virtual agent [9]. All 12 participants correctly identified at least one system function, such as video playback or conversation through questions. Notably, 11 out of 12 participants correctly

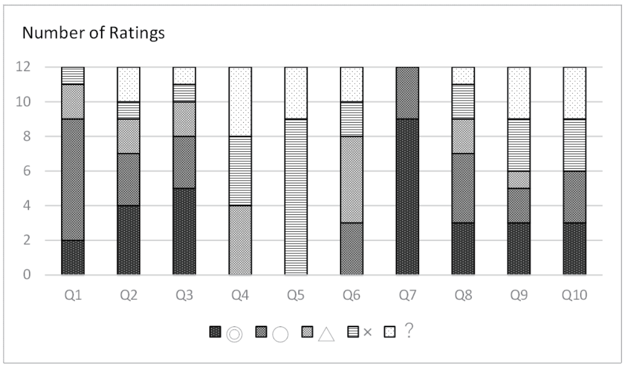


Fig. 4. Breakdown of scoring results for Q3.

identified the scheduling function. However, only 3 out of 12 participants accurately comprehended the overall functionality of the virtual agent in Q3. Many of the incorrect answers indicated that participants perceived the system primarily as a video playback and schedule management system, indicating a need for an understanding of the multiple functions of the system and the inability to infer an overview by synthesizing these functions.

E. Discussion

The analysis of RQs 1, 2, and 3 confirms that project inferences are contingent upon two factors. The first is the extent to which the corpus, $corpus_{Pj}$, accurately reflects the information possessed by the system. In the analysis of RQ1, we provided an example where $corpus_{Pj}$ needed to be improved in information on the project's functionality. As $corpus_{Pj}$ is constructed from class names and component words, it is imperative that the original class names accurately represent the system's characteristics. The second factor is the participant's familiarity with the words present in the corpus. In the analysis of RQ2, we identified instances where the corpus contained words related to the technologies utilized. Participants' knowledge is imperative in correctly interpreting these technologies, resulting in the amount of information obtained from the words varying among individuals [8]. RQ3 also confirms that understanding the system's functions and technologies employed only sometimes results in a correct overview of the project as a whole. Though the system comprises various functions, it is necessary to infer its overall purpose by synthesizing them. To make accurate inferences from the given information, knowledge and experience in software engineering are essential.

F. Further Challenges

One of the challenges associated with the proposed method is structuring the corpus to include the information contained within the classes. Developers differ on how many functions and roles they can include in a single class. If the proposed method is for a class with multiple methods, it is possible to ignore the function and purpose of the class. Therefore, it is expected that $corpus_{Pj}$ can more accurately reflect the roles of the class by obtaining the names of the methods and member variables within the file. Furthermore, the current study did not consider variations in the individual characteristics of participants. The examination of the research question highlighted that the knowledge and experience of the participant impacts the inference process utilizing $corpus_{Pj}$. In future research, it would be beneficial to investigate not only the experience and knowledge of participants but also the specific words they focused on and the information they extracted.

V. Conclusion

In this study, we evaluated the effectiveness of the proposed method from previous research. Specifically, we reevaluated the survey items and examined the information obtained through the corpus from the perspectives of project function, technology employed, and overall system overview. Additionally, we recruited 12 participants and analyzed the results accordingly. The experiment yielded responses from 12 participants. It was determined that the participants comprehension of the project overview relied on two factors: the accuracy of $corpus_{Pj}$ in summarizing the system and the level of knowledge possessed by the participants. Future tasks include incorporating additional information into the corpus by extracting method names and member variables within the classes and structuring the corpus. The study also revealed that the accuracy of a question is contingent upon the participant. Therefore, we intend to investigate the impact of participants' foundational knowledge on the results by documenting the information and techniques they utilized for inference.

Acknowledgment

This research was partially supported by JSPS KAKENHI Grant Numbers JP19H01138, JP20H05706, JP20H04014, JP20K11059, JP22H03699, JP19K02973, Grant-in-Aid for JSPS Research Fellow (No.22J13217), and Tateishi Science and Technology Foundation (C) (No.2207004).

References

[1] V. Lenarduzzi, T. Besker, D. Taibi, A. Martini, and F. A. Fontana, "A systematic literature review on technical debt prioritization: Strategies, processes, factors, and tools," *Journal of Systems and Software*, vol. 171, p. 110827, 2021.

[2] T. McEnery, *Corpus linguistics*. Edinburgh University Press, 2019.

[3] K. Terakawa, S. Chen, and M. Nakamura, "Preliminary study of reasoning existing projects' descriptions based on classname word elements," in *23rd ACIS International Summer Virtual Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD2022-Summer)*, 2022.

[4] B. Bridgens, M. Powell, G. Farmer, C. Walsh, E. Reed, M. Royapoor, P. Gosling, J. Hall, and O. Heidrich, "Creative upcycling: Reconnecting people, materials and place through making," *Journal of Cleaner Production*, vol. 189, pp. 145–154, 2018.

[5] S. Ducasse and D. Pollet, "Software architecture reconstruction: A process-oriented taxonomy," *IEEE Transactions on Software Engineering*, vol. 35, no. 4, pp. 573–591, 2009.

[6] M. A. Hearst, E. Pedersen, L. Patil, E. Lee, P. Laskowski, and S. Franconeri, "An evaluation of semantically grouped word cloud designs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 9, pp. 2748–2761, 2020.

[7] "Android sdk documentation," https://developer.android.com/, (Accessed on 1/30/2023).

[8] P. Spachos and K. N. Plataniotis, "Ble beacons for indoor positioning at an interactive iot-based smart museum," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3483–3493, 2020.

[9] S. Tokunaga, K. Tamamizu, S. Saiki, M. Nakamura, and K. Yasuda, "Virtualcaregiver: Personalized smart elderly care," *International Journal of Software Innovation (IJSI)*, vol. 5, no. 1, pp. 30–43, 2017.