Journal of
Information Systems Engineering
and Business Intelligence

Vol.10, No.1, February 2024
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Analyzing Variances in User Story Characteristics: A Comparative Study of Stakeholders with Diverse Domain and Technical Knowledge in Software Requirements Elicitation

**Ersalina Trisnawati [1)] , Indra Kharisma Raharjana [2)*], Taufik [3)] ,
Ahmad Hoirul Basori [4)], Nouf Atiahallah Alghanmi [5)] , Andi Besse Firdausiah Mansur [6)]**

[1)2)3)] *Information Systems, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia*
[1)]ersalina.trisnawati-2017@fst.unair.ac.id, [2)]indra.kharisma@fst.unair.ac.id, [3)]taufik@fst.unair.ac.id

[4)5)6)] *Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Rabigh, Saudi Arabia*
[4)]abasori@kau.edu.sa, [5)]naalganmy@kau.edu.sa, [6)]abmansur@kau.edu.sa

Abstract

Background: In Agile software development, an essential initial stage is eliciting software requirements. This process engages stakeholders to achieve comprehensive results. However, a common issue is the variance in domain and technical knowledge among stakeholders, potentially impacting the quality of software requirements elicitation.
Objective: Understanding the characteristics of user stories produced by stakeholders becomes crucial, particularly considering the differences in domain and technical knowledge. This study aims to compare the characteristics of user stories generated by stakeholders with varying backgrounds in domain and technical expertise.
Methods: The initial step involves categorizing respondents into distinct stakeholder groups. Three stakeholders are involved in this study, constituting a combination of those with high and low technical and domain knowledge. Subsequently, data collection of user stories is conducted across various case studies. Finally, the acquired user stories are analyzed for further insights.
Results: The analysis reveals variations in user stories generated by the three stakeholder categories across the three case studies. Stakeholders with domain knowledge tend to focus on 'what' aspects with task elements and 'why' aspects with hard-goal elements. Meanwhile, technical knowledge crafts user stories with capability elements in the 'what' aspect. Utilizing the QUS framework, it is evident that technical knowledge consistently produces a higher number of high-quality user stories across all quality categories,
Conclusion: The contribution offered by this study lies in determining the distinct characteristics of user stories produced by different types of stakeholders, focusing on disparities in domain and technical knowledge. The study highlights the comparison of various characteristics of user story elements, such as hard-goals, soft-goals, tasks, or capabilities, and assesses the quality of user stories based on the user story framework. Additionally, it endorse the importance of process innovation in shaping the requirements gathering process and subsequently influencing the quality of user stories.

Keywords: User story, Agile Software Development, Requirements Elicitation, Stakeholder, Domain Knowledge, Process Innovation

Article history: Received 16 January 2024, first decision 1 February 2024, accepted 13 February 2024, available online 28 February 2024

I. INTRODUCTION

The user story format comprises three aspects of requirements: the aspect of who (identifying who wants the functionality), the aspect of what (what functionalities stakeholders want from the system), and the aspect of why (clarifying the purpose behind stakeholders needing this function) [1], [2]. User stories are among the most widely used artifacts in Agile Software Development [3]. They are easy to comprehend due to their concise and straightforward writing format, making them well-suited for writing requirements elicitation [4]. The involvement of stakeholders in the requirements elicitation process is crucial as it significantly influences the outcomes of engineering or software development [5].

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.10.1.110-125

Differences in stakeholders' knowledge can impact the quality of software requirements. Efforts to obtain high-quality requirements can be undertaken by engaging stakeholders in activities such as brainstorming, observation, user interviews, and other relevant methods [6]. It is imperative to understand the characteristics of user stories produced by stakeholders, particularly variations in domain and technical knowledge [7]. Domain knowledge pertains to understanding business processes related to the application under development, while technical knowledge refers to proficiency in software development [8]. Requirements elicitation is an early process in the engineering or software requirements development stage [9], [10]. Incomplete requirements can significantly impact the escalation of maintenance costs and development efforts [11]. To achieve effective elicitation results, software projects typically need to engage numerous stakeholders [12],[13], [14], [15]. Experts possessing domain and software knowledge are crucial in generating high-quality requirements [7]. Conversely, requirements analysts with extensive knowledge of software technology but lacking comprehension of the problem domain may produce low-quality requirements [16]. Therefore, stakeholders' background knowledge can influence the quality of the results in the requirements elicitation process.

This study aims to compare the characteristics of user stories generated by stakeholders with varying backgrounds in domain and technical knowledge. Three types of stakeholders are involved in this study, representing a combination of those with high and low technical and domain knowledge. The research output includes an analysis of the differences in user stories produced by these three types of stakeholders, considering the level of user story elements (hard-goal, soft-goal, task, or capability) [17], and the quality of user stories based on the quality user story framework [18]. The contribution of this study lies in determining the characteristics of user stories produced by different types of stakeholders, categorized by differences in domain and technical knowledge. The analysis encompasses the number of user stories produced, the types of user story elements (hard-goal, soft-goal, task, or capabilities), and the overall quality of the user stories.

II. Literature Review

A user story is a sentence that describes functional requirements in software development as required by users or stakeholders [19]. The user story format encompasses three aspects of requirements: the aspect of who (identifying the entity desiring the functionality), the aspect of what (listing the functionalities stakeholders seek from the system), and the aspect of why (clarifying the purpose behind stakeholders needing this function) [1]. User stories are composed in a semi-structured language following the format: "As <user type>, I want <goal> so <reason>" [20].

User stories also serve to communicate about the system among individuals involved in software development, regardless of their background in information technology knowledge. The user story component is illustrated in Fig. 1. The classification of user story elements used in this study aligns with the definition provided by [17].

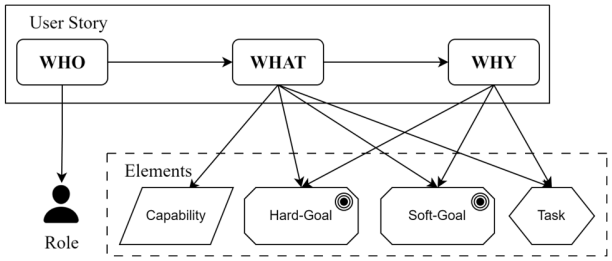


Fig. 1 Component of user story, Adapted from [17]

- Task: User story elements that determine a specific approach to achieving a goal or addressing a particular aspect of why.
- Capability: User story elements representing a user taking action to fulfill a purpose or address an aspect of why, subject to certain conditions and events.
- Hard-goals: User story elements describing the conditions or criteria that stakeholders aim to achieve.
- Soft-goals: User story elements delineating the conditions or criteria the actor desires. However, clear criteria for determining the fulfillment of these conditions are not specified.

The stakeholder categories shown in Fig. 2 are divided into three categories. The first category is Stakeholders Category I (High Domain Knowledge and Low Technical Knowledge). Stakeholders in this category have knowledge related to the problem domain or the software case study to be developed. Additionally, stakeholders in this category do not have knowledge of software engineering or similar fields. The second category is Stakeholders Category II

(Low Domain Knowledge and Low Technical Knowledge). Stakeholders in this category lack knowledge related to the problem domain or the software case study to be developed and also lack knowledge of software engineering or similar fields. The third category is Stakeholders Category III (Low Domain Knowledge and High Technical Knowledge). Stakeholders in this category lack knowledge related to the problem domain or the software case study to be developed. However, they possess knowledge of software engineering or similar fields.

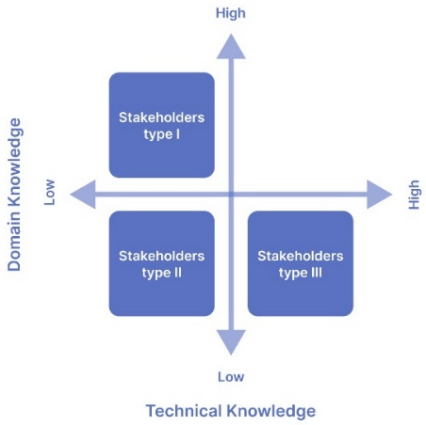


Fig. 2 Types of stakeholders

To enhance the quality of user stories, [18], [21], [22] have formulated criteria for assessing their quality. The QUS Framework categorizes these criteria into three groups: (1) Syntactic quality: Criteria about the textual structure of the user story without considering its meaning; (2) Semantic quality: Criteria concerning the relationship, meaning, or parts of the user story text, and (3) Pragmatic quality: Criteria that take into account the subjective interpretation of stakeholders regarding the user story text, beyond syntax and semantics. For a detailed criteria breakdown, refer to Table 1.

TABLE 1
QUALITY USER STORY FRAMEWORK [18]

Criterion	Description	Individual/set
Syntactic		
Well-formed	A user story includes at least a role and a means	Individual
Atomic	A user story expresses a requirement for exactly one feature	Individual
Minimal	A user story contains nothing more than role, means, and ends	Individual
Semantic		
Conceptually sound	The means express a feature, and the ends describe a rationale	Individual
Problem-oriented	A user story only specifies the problem, not the solution to it	Individual
Unambiguous	A user story avoids terms or abstractions that lead to multiple interpretations	Individual
Conflict-free	A user story should not be inconsistent with any other user story	Set
Pragmatic		
Full-sentence	A user story is a well-formed full sentence	Individual
Estimatable	A story does not denote a coarse-grained requirement that is difficult to plan and prioritize	Individual
Unique	Every user story is unique, duplicates are avoided	Set
Uniform	All user stories in a specification employ the same template	Set
Independent	The user story is self-contained and has no inherent dependencies on other stories	Set
Complete	Implementing a set of user stories creates a feature-complete application, no steps are missing	Set

III. METHODS

The research procedure outlined in Fig. 3 involves several key steps. Initially, the study begins by defining the problem as a case study of domain issues intended for stakeholders to craft user stories during the data collection stage. Subsequently, the process entails determining the respondents or stakeholders for the research, potentially involving classifying respondents into specific groups, as detailed earlier (Stakeholders Type I, II, III). The data collection phase then ensues, involving collecting user story data from identified stakeholders drawn from three distinct case studies: Applications related to Zero Waste, Wastewater Treatment, and Covid-19 Medical Waste. The analysis of user stories encompasses sub-steps such as evaluating quantity, classifying elements within the stories (e.g., hard-goals, soft-goals, task, or capabilities), and assessing quality based on the QUS Framework, considering

criteria like syntactic, semantic, and pragmatic aspects. Collectively, these steps form a comprehensive research procedure designed to identify, collect, and analyze user story data within the study's context.

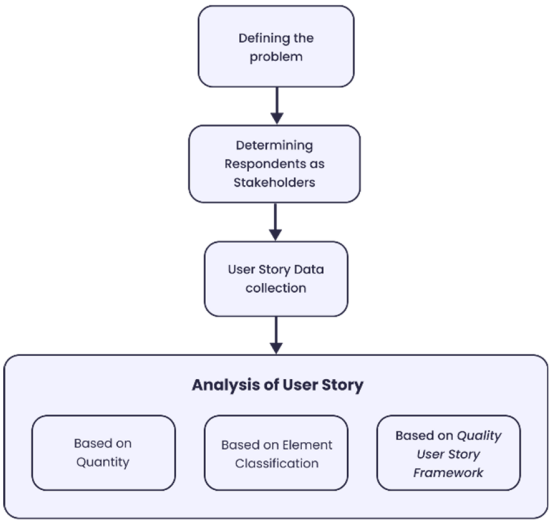


Fig. 3 Research procedure

A. Defining the Problem

The problem in this research is defined as a case study of domain issues used by stakeholders in composing user stories during the user story data collection stage. This study incorporates three case studies to ensure a comprehensive examination: Applications related to Zero Waste, Wastewater Treatment, and Covid-19 Medical Waste.

B. Determining Respondents as Stakeholders

Before collecting user stories, it is essential to identify stakeholders based on their level of domain and technical knowledge background. The division of stakeholder groups is illustrated in Fig. 2. The stakeholder groups are:

1) Stakeholders Type I – HDLT (High Domain Knowledge and Low Technical Knowledge): In this category, respondents are selected based on the criteria of being students with study programs in Environmental Engineering or Biology. This selection is grounded in the understanding that students from these departments possess sufficient domain knowledge related to case studies in application development.

2) Stakeholders Type II - LDLT (Low Domain Knowledge and Low Technical Knowledge): For this category, respondents are chosen from study programs unrelated to information technology/computers/information systems and environmental engineering.

3) Stakeholders Type III – LDHT (Low Domain Knowledge and High Technical Knowledge): In this category, respondents are selected based on the criteria of being final-year students in the Information Systems study program or alums working in software development.

Respondents were chosen according to the categories of stakeholders I-III, with each stakeholder type represented by a minimum of 5 respondents for each case study. The selection of respondents was guided by several criteria outlined in Table 2.

Determining respondents in this study utilized a purposive sampling technique, wherein the researcher selected respondents who met the criteria outlined in Table 2. Respondent data for this study were obtained by completing a questionnaire via Google Form. The questionnaire included inquiries about educational backgrounds and terms related to case studies and software development. This step was undertaken to ascertain whether the respondents fulfilled the criteria essential for the research. The distribution and collection of questionnaires occurred from May 10, 2020, to May 24, 2020. Following the questionnaire collection, 49 responses were gathered, and 45 valid respondents were identified as meeting the research criteria. These respondents were categorized into each stakeholder group in each case study. However, data from four respondents were deemed invalid because they could not be contacted to complete the user story data collection questionnaire. the respondents participating in the questionnaire information from each stakeholder category can also be seen in Table 2.

TABLE 2
RESPONDENT CRITERIA AND PARTICIPATION

Case study	Stakeholder category	Criteria	Respomdemts participation
zero waste	Type I - HDLT	Stakeholders understand the zero waste concept but do not understand the software development process.	5 undergraduate students majoring in Environmental Engineering
	Type II - LDLT	Stakeholders do not understand the zero waste concept and the software development process.	2 management students, 1 nursing student, 1 industrial engineering student, and 1 accounting student
	Type III – LDHT	Stakeholders do not understand the zero waste concept but understand the software development process.	5 undergraduate students majoring in Information Systems
Wastewater Treatment	Type I - HDLT	Stakeholders understand the wastewater treatment concept but not the software development process.	4 undergraduate students majoring in Environmental Engineering and 1 industrial chemical engineering student
	Type II - LDLT	Stakeholders do not understand the concept of wastewater treatment and the software development process.	1 management student, 1 nursing student, 1 interior design student, and 2 medical students
	Type III – LDHT	Stakeholders do not understand the wastewater treatment concept but understand the software development process.	5 undergraduate students majoring in Information Systems
Covid-19 Medical Waste	Type I - HDLT	Stakeholders understand the concept of Covid-19 Medical Waste but not the software development process.	4 undergraduate students majoring in Environmental Engineering and 1 pharmacy education student
	Type II - LDLT	Stakeholders do not understand Covid-19 Medical Waste and the software development process.	2 psychology students, 1 law student, 1 industrial electrical engineering student, and 1 urban and regional planning student
	Type III – LDHT	Stakeholders do not understand the concept of Covid-19 Medical Waste but understand the software development process.	5 undergraduate students majoring in Information Systems

C. User Story Data Collection

The user story collection process involves interviews with respondents, during which they are requested to compose user stories. Respondents in this study will function as representatives of stakeholders. The selection of respondents comprised students or graduates from universities. A questionnaire will be administered first to assess each respondent's domain and technical knowledge. The questionnaire includes various inquiries related to case studies and software development. Each stakeholder is expected to generate 5 to 15 user stories based on the provided case studies. Before composing user stories, stakeholders will receive an explanation covering the research objectives, user stories, and case studies.

User story data is collected by completing an online questionnaire through Google Forms. Each respondent fills out a questionnaire tailored to their domain and technical knowledge. In administering the questionnaire, participants diligently obtain informed consent, signifying their voluntary agreement to partake in the study. Participants are assured that all data about personal information shared in the questionnaire will be treated with confidentiality.

D. Analysis Of User Story

The analysis will involve a comparison of the characteristics of the user stories generated by each stakeholder in every case study. A group of system analysts will conduct the analysis. The systems analysts participating in this study will be final-year students of the Information Systems study program at Universitas Airlangga, with the criteria of having completed software engineering-related courses. This analysis prioritizes technical knowledge to classify types of user stories and evaluate the quality of user stories.

The kappa coefficient value will be examined using the SPSS application (version 25) to measure the agreement among analysts regarding the assessment of each user story. The kappa coefficient value can indicate the strength of agreement between two observers or analysts [23]. The levels of agreement for the kappa value can be defined as outlined in Table 3 [24].

Analysis of user stories is divided into 3 categories: 1) analysis based on quantity, 2) analysis based on element classification, and 3) analysis based on The Quality Of User Story Framework. The basis of this analysis is to obtain user story characteristics and variances. We argue that this analysis can be obtained by determining the quantity produced by stakeholders, the type of classification of user stories, and the quality of the user stories. We use the element classification definition from the Component of user story [17] and add analysis based on quantity. We also analyze the Quality of the resulting User Stories using the Quality of User Story Framework [18].

1) Analysis Based on Quantity

User stories generated by each stakeholder in each case study will be analyzed by quantifying their numbers. Subsequently, the average number of user stories generated by each category of stakeholders will be computed for each case study. The minimum and maximum number of user stories generated will also be determined. This analysis aims to ascertain the number of user stories that each category of stakeholders with varying domain and technical knowledge can generate.

2) Analysis Based on Element Classification

Each user story will undergo classification based on the type of element, specifically hard goals, soft goals, tasks, and capabilities. The classification of each user story will be established according to the elements marked by the analyst, identifying user story elements based on the "what" and "why" aspects. In the "what" aspect, it can be a task/capability/hard goal/soft goal, while in the "why" aspect, it can be a task/hard goal/soft goal. Each analyst will assign a value of 1 to the appropriate element and 0 to other elements unsuitable for each user story in each case study. Subsequently, for each category of stakeholders in each case study, the percentage of each element produced will be calculated. The analysis results will be compared to determine the characteristics of user stories generated by each category of stakeholders with diverse domain and technical knowledge.

3) Analysis Based on the Quality of User Story Framework

The quality of each user story will be measured using 13 criteria based on the Quality User Story (QUS) Framework [18]. The quality of each user story will be determined based on the assessment results of multiple analysts. Each analyst will assess each user story, assigning a value of 1 for "yes" or meeting the criteria and 0 for "no" or not meeting the criteria. Subsequently, for each category of stakeholders in each case study, the average number of user story quality criteria will be computed. The analysis results will be compared to evaluate the quality of user stories produced by each category of stakeholders with distinct domain and technical knowledge.

TABLE 3
LEVEL AGREEMENT

Kappa value	Level agreement
k < 0,20	*Poor*
0,21 – 0,40	*fair*
0,41 – 0,60	*Moderate*
0,61 – 0,80	*Good*
K > 0,81	*Very good*

IV. RESULTS

Based on the aim of this research, which is to understand the characteristics of user stories produced by stakeholders, it becomes important, especially considering the differences in domain and technical knowledge. This study aims to compare the characteristics of user stories produced by stakeholders with various domain backgrounds and technical expertise. User story analysis is conducted based on research analysis procedures, analysis based on quantity, analysis based on element classification, and analysis based on quality user story framework.

A. Analysis Based on Quantity

The analysis of user story data in this study employed descriptive research methods, which aim to depict or describe a condition as it is [27]. This research involved comparing the minimum, maximum, and average number of user stories produced by each category of stakeholders. The summary of the number of user stories generated by each stakeholder category in each case study is presented in Table 4.

According to the data presented in Fig. 4(a), all three stakeholder categories exhibit at least 5 user stories. Regarding the average number of user stories, type 3 demonstrates the highest output with seven stories, while types 1 and 2 generate an average of 5.8 user stories. In terms of the maximum number of user stories, type 3 also reaches 10, whereas types 1 and 2 achieve a maximum of 7.

Examining the graph in Fig. 4(b), the three stakeholder categories again produce at least 5 user stories. For the average number of user stories, type 1 yields the highest at 7.2, type 2 makes the least with six stories, and type 3 maintains an average of 6.4. Regarding the maximum number of user stories, both types 1 and 2 reach ten stories, while type 3 attains a maximum of 8.

The graph in Fig. 4(c) shows that the three stakeholder categories generate at least 5 user stories. Regarding the average number of user stories, type 3 leads with 6.4 stories, while type 2 produces the lowest average at 5.4. Type 1 closely follows type 3 with an average of 6 stories. Examining the maximum number of user stories, type 1 produces

the most at 10, type 2 reaches a maximum of 7, and type 3 achieves a maximum of 8, with minimal difference from type 2.

TABLE 4
SUMMARY OF THE ACQUIRED USER STORY DATA.

Stakeholders' category	Case study	Number of User Story					Sum	Min	Mean	Max
	zero waste	5	7	5	5	7	29	5	5.8	7
Type I - HDLT	Wastewater Treatment	10	5	5	6	10	36	5	7.2	10
	Covid-19 Medical Waste	5	5	5	10	5	30	5	6	10
	zero waste	5	7	5	7	5	29	5	5.8	7
Type II - LDLT	Wastewater Treatment	5	5	5	10	5	30	5	6	10
	Covid-19 Medical Waste	5	5	5	7	5	27	5	5.4	7
	zero waste	10	5	7	5	8	35	5	7	10
Type III – LDHT	Wastewater Treatment	5	8	7	5	7	32	5	6.4	8
	Covid-19 Medical Waste	5	8	5	7	7	32	5	6.4	8

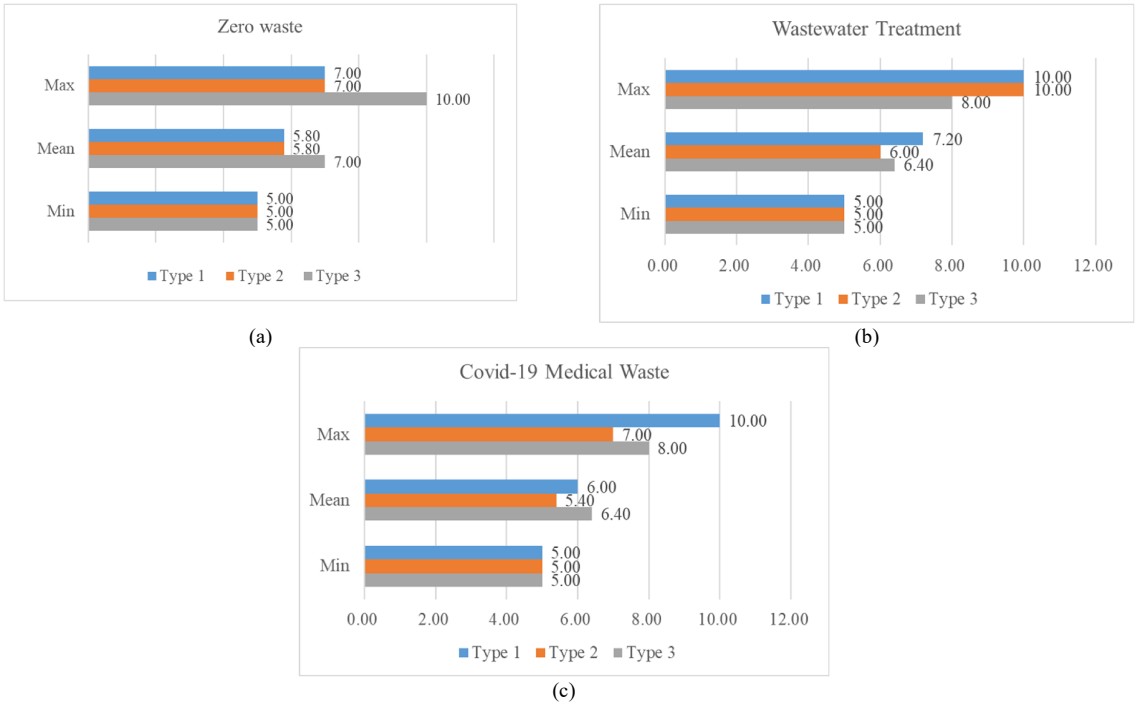


Fig. 4 The figures provide graphical representations of the user story data, showcasing the quantity in the context of (a) a zero waste case, (b) wastewater treatment case studies, and (c) the COVID-19 medical waste case study

B. Analysis Based on Element Classification

The collected user story data is subsequently evaluated by two different analysts, each provided with distinct documents containing the same user story data to ensure an unbiased assessment. After marking user stories, the researchers consolidates the assessment results into a single document. The selection of the highest number of assessments is based on the marked user story results. In cases where user story data is marked differently by analysts, the researchers make the determination regarding the assessment outcomes. To avoid bias towards the results determined by the researchers, confirmation is also sought from two analysts.

The analyst's assessments were evaluated using the kappa coefficient to determine the agreement value between the two analysts. According to the kappa coefficient test results conducted using SPSS (version 25), the kappa values for the three case studies in Table 5 are above 0.81. This value corresponds to a "very good" level of agreement strength, according to Table 3, indicating that the marked user stories based on element classification exhibit a substantial agreement value between the two analysts.

TABLE 5
USER STORY ELEMENT AGREEMENT RESULTS

Case study	Kappa
Zero waste	0.935
Wastewater Treatment	0.907
Covid-19 Medical Waste	0.851

TABLE 6
USER STORY ELEMENT CLASSIFICATION RECAPITULATION

Stakeholders category	Case study	aspect of what				aspect of why		
		Task	**Capability**	**Hard-goal**	**Soft-goal**	**Task**	**Hard-goal**	**Soft-goal**
Type I - HDLT	Zero waste	34%	17%	21%	28%	0%	30%	70%
	Wastewater Treatment	33%	19%	14%	33%	6%	56%	39%
	Covid-19 Medical Waste	43%	17%	17%	23%	5%	55%	40%
Type II - LDLT	Zero waste	14%	14%	24%	48%	5%	64%	32%
	Wastewater Treatment	43%	3%	13%	40%	6%	44%	50%
	Covid-19 Medical Waste	22%	15%	56%	7%	11%	32%	58%
Type III – LDHT	Zero waste	29%	60%	6%	6%	14%	55%	31%
	Wastewater Treatment	19%	69%	13%	0%	42%	37%	21%
	Covid-19 Medical Waste	44%	28%	25%	3%	23%	37%	40%

Following the combination of analysis results, the classification of elements in each stakeholder category for each case study is determined based on the average percentage of each element, using Microsoft Excel. The percentage results are presented in Table 6, categorizing stakeholders as Type I with high domain knowledge and low technical knowledge, Type II with low domain and technical knowledge, and Type III with low domain knowledge and high technical knowledge. In Table 6, highlighted columns indicate the highest percentage of elements in each aspect of user stories generated by each stakeholder category.

In the zero-waste case study, as depicted in Fig. 5(a) and 5(b), Type 1 produces the most user stories with task elements in the 'what' aspect, 34%, and soft goals in the 'why' aspect, 70%. Meanwhile, Type 2 produces the most user stories with soft goal elements in the 'what' aspect, which is 48%. Type 3 makes the most user stories with capability elements in the 'what' aspect, which is 60%. In terms of 'why,' Types 2 and 3 produce the most user stories with hard-goal elements, namely 64% and 55%, respectively.

In the case study of wastewater treatment, as depicted in Fig. 5(c) and 5(d), Type 1 produces the most user stories with task and soft-goal elements in the 'what' aspect, accounting for 33%, and the hard-goal element in the 'why' aspect, which is 56%. Meanwhile, Type 2 generates the most user stories with task elements in the 'what' aspect, amounting to 43%, and soft-goal elements in the 'why' aspect, totaling 50%. Type 3 produces the most user stories with capability elements in the 'what' aspect accounting 69% and task elements in the 'why' aspect is 42%.

In the Covid-19 medical waste case study as seen in Fig. 5(e) and 5(f), Type 1 generated the majority of user stories with task elements in the 'what' aspect, accounting for 43%, and hard-goal elements in the 'why' aspect, with a percentage of 55%. Meanwhile, Type 2 produced the most user stories with hard-goal elements in the 'what' aspect, totaling 56%. Type 3 produced the highest number of user stories with a task element in the 'what' aspect, amounting to 44%. In terms of 'why', Types 2 and 3 generated the most user stories with soft-goal elements, namely 58% and 40%.

C. Analysis Based on Quality User Story Framework

Two separate analysts evaluate the gathered user story data, each receiving unique documents containing identical user story data to ensure impartial evaluation. Following the marking of user stories, the researchers combine the assessment findings into a unified document. The selection of the predominant assessment is determined by the marked user story outcomes. If there are discrepancies in how analysts mark the user story data, the researchers decide on the

Fig. 5 Chart of the classification of user story elements (a) what aspect of the zero waste case study, (b) why aspect of the zero waste case study, (c) what aspect of the Wastewater Treatment case study, (d) why aspect of the Wastewater Treatment case study

assessment results. To mitigate any bias towards the researchers' determinations, confirmation is additionally sought from two analysts.

The results of the analysts' assessments were measured using the kappa coefficient to determine the agreement value between the two analysts. Based on the kappa coefficient test results using SPSS (version 25), the kappa value for the three case studies in Table 7 is above 0.81. This value indicates a very good level of agreement strength, as outlined in Table 1. Thus, it demonstrates that the results of user stories marked based on the QUS framework can be considered reliable due to the strong agreement value between the two analysts.

TABLE 1
USER STORY QUALITY ANALYSIS AGREEMENT RESULTS

Case study	Kappa
Zero waste	0.964
Wastewater Treatment	0.957
Covid-19 Medical Waste	0.879

After combining the results of the analyst's assessments, the quality of user stories based on the Quality User Story (QUS) framework is analyzed for each stakeholder category in each case study. This is done by calculating the average percentage of each QUS criterion using Microsoft Excel. The summarized results of the percentages are presented in Table 8.

TABLE 8
USER STORY QUALITY RECAPITULATION

Stakeholders category	Case study	syntatic			semantic					pragmatic				
		Well-formed	Atomic	Minimal	Conceptually sound	Problem-oriented	Unambiguous	Conflict-free	Full sentence	Estimatable	Unique	Uniform	Independent	Complete
Type 1	Zero waste	100%	76%	96%	57%	92%	90%	100%	97%	78%	100%	0%	80%	60%
	Wastewater Treatment	100%	63%	53%	45%	87%	72%	100%	92%	64%	100%	80%	60%	40%
	Covid-19 Medical Waste	100%	64%	98%	42%	96%	68%	80%	96%	68%	60%	20%	80%	60%
Type 2	Zero waste	100%	100%	96%	79%	73%	85%	60%	86%	81%	80%	40%	80%	100%
	Wastewater Treatment	100%	70%	58%	54%	100%	64%	100%	76%	68%	80%	20%	100%	0%
	Covid-19 Medical Waste	100%	97%	77%	40%	77%	73%	60%	86%	85%	80%	20%	80%	80%
Type 3	Zero waste	100%	98%	100%	70%	100%	87%	80%	94%	80%	60%	60%	60%	80%
	Wastewater Treatment	100%	85%	54%	51%	100%	96%	100%	93%	98%	60%	100%	80%	80%
	Covid-19 Medical Waste	100%	76%	94%	61%	100%	75%	80%	94%	85%	80%	60%	100%	80%

Based on the graph in Fig. 6(a), the analysis of the quality of user stories in the zero waste case study shows that each category of stakeholders produces user stories with different advantages on each criterion. Type 1 makes high-quality user stories for eight criteria: well-formed, minimal, problem-oriented, unambiguous, full sentence, conflict-free, unique, and independent. However, Type 1 produces low-quality user stories for five criteria: atomic, conceptually sound, estimable, uniform, and complete. Type 2 makes high-quality user stories for nine criteria: well-formed, atomic, minimal, unambiguous, full sentence, estimable, unique, independent, and complete. However, Type 2 produces low-quality user stories for 4 criteria: conceptually sound, problem-oriented, conflict-free, and uniform. Type 3 produces high-quality user stories for eight criteria: well-formed, atomic, minimal, problem-oriented, unambiguous, full sentence, conflict-free, and complete. However, Type 3 produces low-quality user stories for five criteria: conceptually sound, estimable, unique, uniform, and independent.

Based on the graph in Fig. 6(b), the analysis of the quality of user stories in the wastewater treatment case study shows that each category of stakeholders produces user stories with different advantages on each criterion. Type 1 produces high-quality user stories for 6 criteria: well-formed, problem-oriented, full sentence, conflict-free, unique, and uniform. However, Type 1 produces low-quality user stories for 7 criteria: atomic, minimal, conceptually sound, unambiguous, estimable, independent, and complete. Type 2 produces high-quality user stories for 5 criteria: well-formed, problem-oriented, conflict-free, unique, and independent. However, Type 2 produces low-quality user stories for 8 criteria: atomic, minimal, conceptually sound, unambiguous, full sentence, estimable, uniform, and complete.

Type 3 produces high-quality user stories for 10 criteria: well-formed, atomic, problem-oriented, unambiguous, full sentence, estimable, conflict-free, uniform, independent, and complete. However, Type 3 produces user stories with low quality for 3 criteria: minimal, conceptually sound, and unique.

Based on the graph in Fig. 6(c), the analysis of the quality of user stories in the Covid-19 medical waste case study shows that each category of stakeholders produces user stories with different advantages on each criterion. Type 1 produces high-quality user stories for 6 criteria: well-formed, minimal, problem-oriented, full sentence, conflict-free, and independent. However, Type 1 produces user stories with low quality for 7 criteria: atomic, conceptually sound, unambiguous, estimable, unique, uniform, and complete. Type 2 produces high-quality user stories for 7 criteria: well-formed, atomic, full sentence, estimable, unique, independent, and complete. However, Type 2 produces user stories with low quality for 6 criteria: minimal, conceptually sound, problem-oriented, unambiguous, conflict-free, and uniform. Type 3 produces high-quality user stories for 9 criteria: well-formed, minimal, problem-oriented, full sentence, estimable, conflict-free, unique, independent, and complete. However, Type 3 produces user stories with low quality for 4 criteria: atomic, conceptually sound, unambiguous, and uniform.

V. Discussion

Based on the aim of this research, which is to understand the characteristics of user stories produced by stakeholders, it becomes important, especially considering the differences in domain and technical knowledge. This study aims to compare the characteristics of user stories produced by stakeholders with various domain backgrounds and technical expertise. User story analysis is conducted based on research analysis procedures, analysis based on quantity, analysis based on element classification, and analysis based on quality user story framework.

A. Analysis Based on Quantity

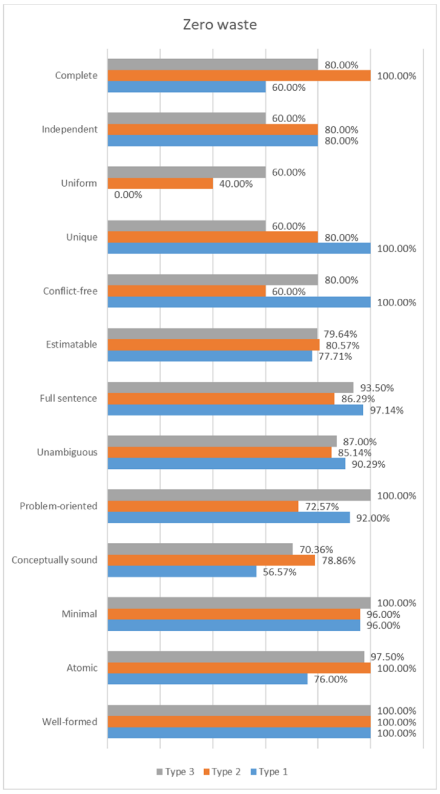
If viewed based on the minimum number of user stories shown in Table 4, the three categories of stakeholders in the three case studies produced the same number of 5 user stories. This amount corresponds to the minimum number of user stories required for each case study. Regarding the average number of user stories, type 1 produces the most user stories in the wastewater treatment case study, with an average of 7.2. Type 2 makes the most user stories in the wastewater treatment case study, with an average of 6, while type 3 produces the most user stories in the zero waste case study, with an average of 7. If viewed based on the maximum number of user stories, the three categories of stakeholders produced the same number, 10, in different case studies—namely, type 1 in the wastewater treatment and Covid-19 medical waste case study, type 2 in the wastewater treatment case study, and type 3 in the zero waste case study.

Based on the user story data in this study, a consistent trend was revealed in the number of stories generated by stakeholders across three different case studies, irrespective of the domain or technical expertise. Each stakeholder category consistently contributes at least five user stories, representing the minimum number of user stories required to be created by respondents. Type 3 stakeholders stand out with the highest average and maximum number of user stories in all cases, indicating a strong contribution to requirements elicitation. Stakeholders with higher technical knowledge demonstrate higher engagement and creativity in formulating requirements.

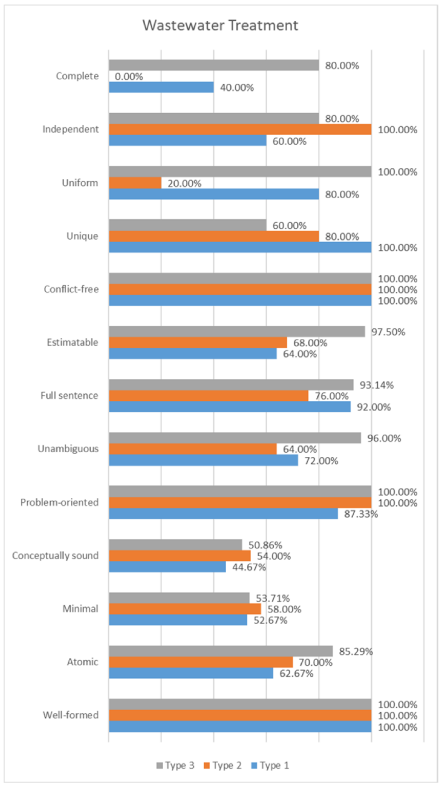
B. Analysis Based on Element Classification

Based on the results of the classification of user story elements generated by stakeholders from the three case studies, type 1 predominantly produces user stories with task elements in the 'what' aspect. However, in the case study of wastewater treatment it makes the exact percentage value for both task and soft-goal elements. Meanwhile, the largest percentage value of user stories generated in the 'why' aspect is the hard-goal element in two of the three case studies, while in the remaining case study, the largest percentage value is for the soft-goal element. Hard-goal and soft-goal elements have a similar meaning, describing the conditions or circumstances to be achieved. The difference is that soft goals do not have clear criteria for achieving their objectives, while hard goals have clearer criteria [17]. It can be observed that type 1 is dominant in writing user stories on the 'what' aspect with task elements and on the 'why' aspect with hard-goal elements.

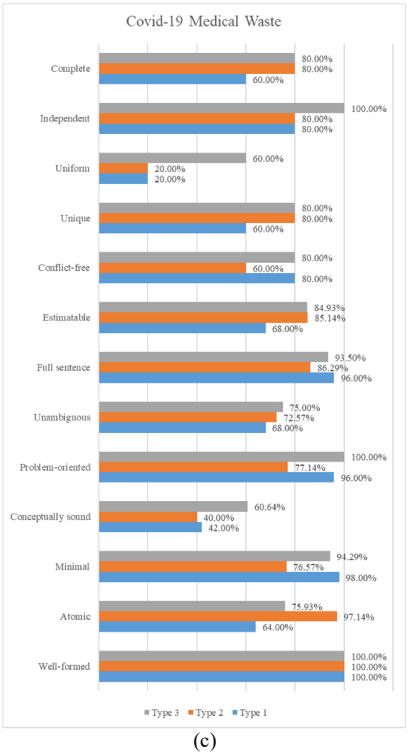
The highest percentage value produced by type 2 has different elements in each case study. In the zero waste case study, type 2 produced the most user stories with soft-goal elements, accounting for 48% shown in the Fig.5(a). In the case study of wastewater treatment shown in the Fig.5(c), the resulting user stories predominantly contain task elements, comprising 43%. However, the percentage of user stories generated with soft-goal elements is also not significantly different, namely 40%. In the Covid-19 medical waste case study , the user stories generated contained



(a)



(b)



(c)

Fig. 6 Graphics of user story quality for (a) Zero waste case study, (b) wastewater treatment case study, (c) Covid-19 medical waste case study

the most hard-goal elements, constituting 56% shown in the Fig.5(e). Meanwhile, the largest percentage value produced by type 2 in the 'why' aspect is the soft-goal element in two of the three case studies, while in the remaining case study, it is the hard-goal element. This shows that type 2 does not dominate specific elements in the 'what' aspect, while type 2 is dominant in producing user stories with soft-goal elements in the' why' aspect.

Type 3 produces user stories with the largest percentage value in the 'what' aspect, featuring capability elements in two of the three case studies, while in another case study, it produces task elements. The elements of task and capability have a similar meaning—namely, a certain, more concrete way to achieve a goal. The difference is that a task is an operational method, while a capability is an action taken by the user [17]. Meanwhile, the largest percentage value generated in the 'why' aspect by type 3 has different elements in each case study. In the zero waste case study, type 3 produces the most user stories with hard-goal elements, constituting 55% shown in the Fig.5(b). In the case study of wastewater treatment, the resulting user story predominantly contains task elements shown in Fig.5(d), accounting for 42%. In the Covid-19 medical waste case study, the resulting user stories predominantly contain soft-goal elements, comprising 40% shown in the Fig.5(e). This shows that type 3 dominantly produces user stories in the 'what' aspect with capability elements, and in the 'why' aspect, type 3 does not have a dominant element.

In the examination of the zero-waste scenario, user stories from Type 1 stakeholders were mainly characterized by task and soft-goal components, whereas Type 2 exhibited a strength in soft goals, and Type 3 showed a focus on capability elements. Analysis of wastewater treatment indicated that Type 1 excelled in both task and soft-goal elements, Type 2 excelled in tasks and soft goals, and Type 3 excelled in capability and task elements. In the Covid-19 medical waste case study, Type 1 demonstrated proficiency in task and hard-goal elements, Type 2 excelled in hard-goal elements, and Type 3 showcased strength in task elements. The investigation into the 'why' aspect of this case study underscored Types 2 and 3 as leaders in generating soft-goal elements.

C. Analysis Based on Quality User Story Framework

If viewed based on the three quality categories of the QUS framework in Table 2, the syntactic category contains three criteria: well-formed, atomic, and minimal. All three categories of stakeholders exhibit high quality on the well-formed criterion in all case studies. On the atomic criterion, types 2 and 3 were superior in two out of the three case studies compared to K1. Regarding the minimal criterion, types 1 and 3 are superior in two out of three case studies compared to type 2. This shows that type 3 dominantly produces user stories with high quality in criteria related to the textual structure of user stories compared to types 1 and 2 [18].

The semantic category contains four criteria: conceptually sound, problem-oriented, unambiguous, and conflict-free. All three categories of stakeholders exhibit low quality on the conceptually sound criterion in all case studies. Types 1 and 3 have high quality on the problem-oriented and conflict-free criteria. Additionally, on the unambiguous criterion, type 3 was superior in two out of the three case studies compared to types 1 and 2. This shows that type 3 produces more user stories with high quality on criteria related to the relationship and meaning of user story texts, while type 2 has low quality in this category [18].

Then, there is the last quality category, namely the pragmatic category, which contains six criteria: full sense, estimable, unique, uniform, independent, and complete. Types 1 and 3 are superior to the full sentence criterion compared to type 2. Additionally, type 3 is also superior in the estimable and complete criteria in all three case studies. Unlike type 2, it excels in unique and independent criteria. Therefore, it is stated that type 2 produces user stories with different ideas or ones not considered by types 1 and 3 [7]. This shows that type 3 produces more user stories with high quality on criteria that consider the subjective interpretation of stakeholders from user story texts, other than syntactic and semantic factors [18].

Using the QUS framework to evaluate user stories provides important insights into their quality in three primary domains: pragmatic, semantic, and syntactic. (1). Syntactic Quality: Every stakeholder ensures that user stories are coherent regardless of classification. While Types 1 and 3 keep things brief and to the point, Types 2 and 3 offer more precise and in-depth information. (2) Semantic Quality: It is critical to comprehend the significance of user stories. Creating theoretically sound user stories is a challenge for all parties. On the other hand, Type 3 excels at providing an unambiguous meaning, whereas Types 1 and 3 excel at handling issues and preventing disputes. (3) Pragmatic Quality: The pragmatic elements of user stories, such as context and completeness, are essential. When it comes to ensuring that user stories make perfect sense, Types 1 and 3 outperform Type 2. Type 2 needs improvement in the areas of distinctiveness and independence, which Type 3 consistently delivers in well-defined, thorough stories.

Type 3 stands out as an exceptional performer, focusing on user stories' independence, clarity, and completeness in addition to accuracy. All stakeholders, especially Type 2, should concentrate on improving conceptual soundness and obtaining a full sense, as these findings offer insightful guidance for progress. In conclusion, consistently producing

high-quality results can be achieved by comprehending and improving user stories' structure, meaning, and practical elements. Stakeholders can collaborate to enhance their user story-generation procedures by utilizing these insights.

D. Study Limitations and Future Research

When requesting user story results from stakeholders, we asked for a minimum of 5 and a maximum of 10 user stories per respondent. In future research, researchers need to consider removing these requirements. In this study, the minimal analysis of user stories yielded results between 5 and 10, which is the minimum and maximum requirement for evaluating user story results. Furthermore, we encountered difficulty in achieving consensus regarding the types of user stories, especially between tasks and capabilities, as well as between hard-goals and soft-goals. To minimize the risk of misunderstanding, we need to align perceptions with analysts on when a user story should be classified as a task, capability, hard-goal, or soft-goal. For future research, it would be easier to adopt a grouping of user stories into two categories: hard/soft goals and task/capability [2].

In this study, we did not analyze the case study results because it requires experts with high domain knowledge and high technical knowledge that are either unavailable or difficult to find. However, we argue that the variances and characteristics in user stories can already be identified based on quantity, element classification, and the quality of user stories.

This study lays the foundation for process improvement in the requirement elicitation process, particularly in the collection of user stories from stakeholders with various domains and technical backgrounds. Further research could explore the dynamics of collaboration among stakeholders from diverse domains and technical backgrounds, deepen the analysis of user story elements, and understand how user story characteristics and quality evolve over time within specific stakeholder groups.

VI. CONCLUSIONS

From the research that has been conducted, several conclusions can be drawn regarding the characteristics of user stories produced by stakeholders with different domains and technical knowledge. The examination of the number of user stories across the three case studies reveals varying outputs among the stakeholder categories. In the wastewater treatment case study, Type 1 led in user story production with an average of 7.2, Type 2 took the lead in the same case study with an average of 6, and Type 3 excelled in the zero-waste case study with an average of 7.

The result of user story analysis based on classification of elements shows that stakeholders with high domain and low technical knowledge are dominant in writing user story on aspect of what with task elements and aspect of why with hard-goal elements. Stakeholders with low domain and technical knowledge do not dominate certain elements in the aspect of what and in the aspect of why producing user stories with soft-goal elements. Stakeholders with low domain knowledge and high technical dominance write user story on aspect of what with capability elements, while on aspect of why do not dominate certain elements.

The evaluation of user story quality, utilizing the QUS framework, indicates that Type 3 stakeholders consistently produce high-quality stories in the syntactic, semantic, and pragmatic categories. However, they exhibit lower quality in specific criteria: conceptually sound, unique, and uniform. Types 1 and 2 also demonstrate lower quality in conceptually sound and uniform criteria, with Type 2 displaying low quality in the semantic category despite excelling in the unique criterion. These findings underscore the need for collaborative efforts among stakeholders to enhance user story generation processes, leveraging insights gained from this comprehensive analysis.

Author Contributions: *Ersalina Trisnawati*: Investigation, Data Curation, Writing - Original Draft. *Indra Kharisma Raharjana*: Conceptualization, Methodology, Writing - Review & Editing, Supervision. *Taufik*: Writing - Review & Editing, Supervision. *Ahmad Hoirul Basori*: Writing - Review & Editing. *Nouf Atiahallah Alghanmi*: Writing - Review & Editing. *Andi Besse Firdausiah Mansur*: Writing - Review & Editing.

All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Fakultas Sains dan Teknologi Universitas Airlangga under Penelitian Unggulan Fakultas Grand number 251/UN3/2022.

Conflicts of Interest: Raharjana and Basori is members of Editorial Teams, but had no role in the decision to publish this article. No other potential conflict of interest relevant to this article was reported.

Data Availability: The data will be made available for sharing upon request.

Informed Consent: Informed Consent was obtained, and a detailed explanation was presented in the Methods section.

Animal Subjects: There were no animal subjects.

ORCID:
Ersalina Trisnawati: -
Indra Kharisma Raharjana: https://orcid.org/0000-0002-0622-3374
Taufik: -
Ahmad Hoirul Basori: https://orcid.org/0000-0001-9684-490X
Nouf Atiahallah Alghanmi: -
Andi Besse Firdausiah Mansur: https://orcid.org/0000-0001-9454-2780

REFERENCES

[1] F. Dalpiaz and S. Brinkkemper, "Agile requirements engineering with user stories," *Proc. - 2018 IEEE 26th Int. Requir. Eng. Conf. RE 2018*, pp. 506–507, 2018, doi: 10.1109/RE.2018.00075.

[2] D. Siahaan, I. K. Raharjana, and C. Fatichah, "User story extraction from natural language for requirements elicitation: Identify software-related information from online news," *Inf. Softw. Technol.*, vol. 158, no. June 2023, p. 107195, Jun. 2023, doi: 10.1016/j.infsof.2023.107195.

[3] I. Inayat, S. Salwah, S. Marczak, M. Daneva, and S. Shamshirband, "A systematic literature review on agile requirements engineering practices and challenges," *Comput. Human Behav.*, vol. 51, pp. 915–929, 2015, doi: 10.1016/j.chb.2014.10.046.

[4] I. K. Raharjana, D. Siahaan, and C. Fatichah, "User Story Extraction from Online News for Software Requirements Elicitation: A Conceptual Model," *JCSSE 2019 - 16th Int. Jt. Conf. Comput. Sci. Softw. Eng. Knowl. Evol. Towar. Singul. Man-Machine Intell.*, pp. 342–347, Jul. 2019, doi: 10.1109/JCSSE.2019.8864199.

[5] M. J. Ryan, "The Role of Stakeholders in Requirements Elicitation," *INCOSE Int. Symp.*, vol. 24, no. 1, pp. 16–26, Jul. 2014, doi: 10.1002/j.2334-5837.2014.tb03131.x.

[6] M. R. Dewi, I. Kharisma Raharjana, D. Siahaan, and C. Fatichah, "Software Requirement-Related Information Extraction from Online News using Domain Specificity for Requirements Elicitation: How the system analyst can get software requirements without constrained by time and stakeholder availability," in *2021 10th International Conference on Software and Computer Applications*, Feb. 2021, pp. 81–87, doi: 10.1145/3457784.3457796.

[7] A. Niknafs and D. Berry, "The impact of domain knowledge on the effectiveness of requirements engineering activities," *Empir. Softw. Eng.*, vol. 22, no. 1, pp. 80–133, 2017, doi: 10.1007/s10664-015-9416-2.

[8] A. M. Aranda, O. Dieste, and N. Juristo, "Effect of Domain Knowledge on Elicitation Effectiveness: An Internally Replicated Controlled Experiment," *IEEE Trans. Softw. Eng.*, vol. 42, no. 5, pp. 427–451, 2016, doi: 10.1109/TSE.2015.2494588.

[9] K. Pohl and C. Rupp, *Requirements Engineering Fundamentals*, 2nd Editio. Rocky Nook Inc., 2015.

[10] H. Dar, M. I. Lali, H. Ashraf, M. Ramzan, T. Amjad, and B. Shahzad, "A systematic study on software requirements elicitation techniques and its challenges in mobile application development," *IEEE Access*, vol. 6, pp. 63859–63867, 2018, doi: 10.1109/ACCESS.2018.2874981.

[11] S. Sharma and S. K. Pandey, "Requirements elicitation: Issues and challenges," in *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, Mar. 2014, pp. 151–155, doi: 10.1109/IndiaCom.2014.6828119.

[12] C. Pacheco and I. Garcia, "A systematic literature review of stakeholder identification methods in requirements elicitation," *J. Syst. Softw.*, vol. 85, no. 9, pp. 2171–2181, 2012, doi: 10.1016/j.jss.2012.04.075.

[13] D. A. P. Sari *et al.*, "Crowdsourcing as a tool to elicit software requirements," *AIP Conf. Proc.*, vol. 2329, no. 1, p. 050001, 2021, doi: 10.1063/5.0042134.

[14] S. L. Lim and A. Finkelstein, "StakeRare: Using social networks and collaborative filtering for large-scale requirements elicitation," *IEEE Trans. Softw. Eng.*, vol. 38, no. 3, pp. 707–735, 2012, doi: 10.1109/TSE.2011.36.

[15] H. Wulandari and T. Raharjo, "Systematic Literature and Expert Review of Agile Methodology Usage in Business Intelligence Projects," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 9, no. 2, pp. 214–227, 2023, doi: 10.20473/jisebi.9.2.214-227.

[16] H. Kaiya and M. Saeki, "Using Domain Ontology as Domain Knowledge for Requirements Elicitation," in *14th IEEE International Requirements Engineering Conference (RE'06)*, Sep. 2006, pp. 189–198, doi: 10.1109/RE.2006.72.

[17] Y. Wautelet, S. Heng, M. Kolp, I. Mirbel, and S. Poelmans, "Building a rationale diagram for evaluating user story sets," in *2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*, Jun. 2016, vol. 2016-Augus, pp. 1–12, doi: 10.1109/RCIS.2016.7549299.

[18] G. Lucassen, F. Dalpiaz, J. M. E. M. van der Werf, and S. Brinkkemper, "Improving agile requirements: the Quality User Story framework and tool," *Requir. Eng.*, vol. 21, no. 3, pp. 383–403, Sep. 2016, doi: 10.1007/s00766-016-0250-x.

[19] Y. Wautelet, S. Heng, D. Hintea, M. Kolp, and S. Poelmans, "Bridging user story sets with the use case model," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9975 LNCS, pp. 127–138, doi: 10.1007/978-3-319-47717-6_11.

[20] C. Thamrongchote and W. Vatanawood, "Business process ontology for defining user story," *2016 IEEE/ACIS 15th Int. Conf. Comput. Inf. Sci. ICIS 2016 - Proc.*, pp. 3–6, 2016, doi: 10.1109/ICIS.2016.7550829.

[21] G. Lucassen, F. Dalpiaz, J. M. E. M. Van Der Werf, and S. Brinkkemper, "Forging high-quality User Stories: Towards a discipline for Agile Requirements," in *2015 IEEE 23rd International Requirements Engineering Conference, RE 2015 - Proceedings*, 2015, pp. 126–135, doi: 10.1109/RE.2015.7320415.

[22] F. Dalpiaz, I. van der Schalk, and G. Lucassen, "Pinpointing Ambiguity and Incompleteness in Requirements Engineering via Information Visualization and NLP," in *Requirements Engineering: Foundation for Software Quality. REFSQ 2018. Lecture Notes in Computer Science*, vol. 10753 LNCS, no. March, E. Kamsties, J. Horkoff, and F. Dalpiaz, Eds. Springer International Publishing, 2018, pp. 119–135.

[23] E. L. MacQuillan, J. Ford, and K. Baird, "Clinical Competence Assessed Using Simulation: Development of a Standardized Tool to Assess Nutrition-Focused Physical Exam Skill Competence for Registered Dietitian Nutritionists," *J. Nutr. Educ. Behav.*, vol. 53, no. 2, pp. 174–182, 2021, doi: 10.1016/j.jneb.2020.10.014.

[24] Laerd Statistics, "Fleiss' kappa using SPSS Statistics," *Statistical tutorials and software guides*, 2019. https://statistics.laerd.com/spss-tutorials/fleiss-kappa-in-spss-statistics.php (accessed Feb. 10, 2024).