Journal of
Information Systems Engineering
and Business Intelligence

Vol.11, No.3, October 2025
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Generating User Personas for Eliciting Requirements Using Online News Data

Halim Wildan Awalurahman [1], Indra Kharisma Raharjana [2]*, Kartono Kartono [3], Shukor Sanim Mohd Fauzi [4]

[1)2)] *Center for Information Systems Engineering, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia*
[1)]halim.wildan.awalurahman-2019@fst.unair.ac.id, [2)]indra.kharisma@fst.unair.ac.id

[2)3)]*Department of Information Systems, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia*
[2)]indra.kharisma@fst.unair.ac.id, [3)]kartono @fst.unair.ac.id

[4)] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perlis Branch, Arau Campus, Arau, Perlis 02600, Malaysia*
[4)] shukorsanim @uitm.edu.my

Abstract

Background: In software development, creating user personas remains challenging despite their recognized value. Cost, adaptability, and data scarcity present obstacles in designing these critical personas. A new perspective and process innovation for generating user personas is essential to overcome this hurdle.
Objective: This study introduces a method for extracting user persona attributes, including names, occupations, workplaces, and goals.
Methods: A framework for extracting user persona information from online news sources is created. Our method employs a comprehensive SpaCy processing pipeline, incorporating NER, SpaCy rule-based matching, and phrase matching.
Results: The evaluation results showcase promising performance metrics, with an average recall value of 0.700, precision of 0.402, and F1-score of 0.506.
Conclusion: This study demonstrates the feasibility of extracting user persona attributes from online news data. Future research could focus on enhancing the method's performance, investigating its effectiveness in creating relationships, and ensuring that the generated user personas accurately reflect the news text data.

Keywords: Process innovation, Natural Language Processing, Online News, Software Development, User Persona

Article history: Received 23 August 2025, first decision 24 September 2025, accepted 13 October 2025, available online 28 October 2025

I. INTRODUCTION

Elicitation of requirements is a key milestone in software development [1], [2], [3]. Active user involvement throughout the software development lifecycle is imperative to ensure that the resulting software aligns seamlessly with their needs and expectations [4], [5], [6]. User personas are among the essential artifacts in software development that bridge the gap between users and their requirements [7], [8]. A user persona is a fictional representation of a group of users that illustrates key user characteristics, needs, and behaviors, supporting development teams in understanding and prioritizing product development [9]. A user persona tells us what users are like and what they want to do [10], [11]. It helps us identify potential problems that may arise when designing the software, allowing us to understand what the user wants, what the system should do, and what should be excluded [12], [13]. A user persona encompasses several aspects, including goals, challenges, psychographics, demographics, and, if available, pictures [14]. In this situation, user personas become essential tools for aligning software development efforts with user-centric viewpoints, increasing the likelihood of producing software that meets user expectations.

The user persona has been used in modern software development phases, from early communication with stakeholders to development and end-user training [15]. They serve as a valuable tool for developers and design teams

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.11.3.407-419

to understand user preferences or goals [16]. A user persona is advantageous for professionals, such as UI/UX designers and system analysts, who engage with user needs [17].

Although it is beneficial and functional in software development, creating user personas persists with numerous challenges. Developing user personas can be costly, susceptible to frequent changes, and hindered by limited data availability [18], [19], [20]. Typically, researchers rely on interviews to construct user personas, a method that poses significant challenges due to time and location constraints and the need to engage with users [21]. Although qualitative methods provide rich contextual understanding, they lack scalability and reproducibility when applied to large-scale or rapidly evolving digital domains [15]. Therefore, a new approach to creating a user persona is necessary to address the challenges [5], [19].

Several studies have extracted information, including features and user stories, from online news sources [22], [23], [24]. Notably, a study from [25] demonstrated the feasibility of extracting quotations from online news. Meanwhile, [23], [24] provided evidence supporting the creation of user stories from online news—a concept closely resembling user personas, given its focus on actors and goals. These investigations have unveiled promising prospects for deriving user persona attributes from online news sources.

We refer to [25], which can extract direct and indirect information from online news text, to support the use of online news.Click or tap here to enter text. This study demonstrated the potential of using online news to identify quotations that mention actors, entities, or goals. The findings of this study align with our goal of identifying aspects of a user persona, particularly in terms of names and goals. Referring to our output, we refer to [26], which explored how to create a better persona using affordances. Although this study still employs qualitative methods, such as interviews, we believe that it contains sufficient information to determine that our output of user personas will include names, jobs, workplaces, and goals.

News reports often feature real-life cases containing information about people expressing their goals or desires. In some instances where the news is relevant to software development, this information can be invaluable in eliciting needs, particularly in understanding the domain. Using online news as a source of requirements can minimize constraints on software development projects, such as time and limited stakeholders' access [27], [28]. The approach of asking users direct questions may be less effective because users experiencing specific problems cannot automatically understand the solutions. Therefore, a study from [30] suggests using pretend user personas as hypothetical archetypes of actual users to view solutions holistically. By creating user personas based on user taxonomy, one can identify users' general characteristics in detail without getting bogged down in the individual desires and behavioral anomalies of real users. Online news emerges as one of the most interesting sources for considering user personas.

In software development, creating user personas is essential for understanding user needs [29]. Traditionally, personas are constructed from user observations gathered through interviews or surveys; however, this process is often time-consuming [30]. This study proposes a new approach to persona creation that leverages online news data as an additional source of insight. While traditional methods focus on individual behavior, online news data provide valuable context about broader user trends and challenges. Although online news may not directly describe end users in the software development process, it often contains domain-relevant issues and stakeholder opinions that can reflect potential user needs and goals. These extracted insights serve as the initial representations of the user personas. This study develops a method for generating user personas from online news by extracting key attributes, including names, occupations, workplaces, and goals. The contribution of this study lies in demonstrating the feasibility of this automated approach.

II. METHODS

This study aims to develop a method for extracting user personas, including names, job titles, workplaces, and goals. We have adapted the methods outlined in [23], which were originally intended for generating user stories, and have made specific adaptations align with our objectives. This method's selection is reinforced by its success in extracting user story aspects (who, what, and why) from online news, which closely aligns with this study's goal of extracting user persona aspects from online news. Given the diversity of persona attributes, this study focuses on extracting a subset of user personas, such as names, jobs, workplaces, and goals, that are feasible to capture automatically from online news using NER and rule-based matching. This study examines the process of extracting information from online news to construct user personas and explores how this can be accomplished. The proposed method is illustrated in Fig. 1. The research progresses through the following specific phases: user persona formulation, data collection, information extraction, and evaluation. The NLP processing in this study was conducted using SpaCy, a widely used open-source library for NLP tasks [31].

Online news articles provide contextual information and stakeholder perspectives related to real-world issues. Statements from founders, organizations, or domain experts are interpreted as stakeholder viewpoints that can represent user expectations or requirements. While it may seem unconventional to base personas on online news data, understanding its value as a supplementary source of insight is crucial. Online news data provides a quick perspective on trends and challenges. By analyzing news articles related to specific topics or industries, we can uncover patterns and insights that enhance persona development. Relevant news sources and topics were carefully selected and considered to address potential biases and coverage limitations. To mitigate these constraints, we carefully choose relevant news sources and topics for our study. Additionally, we propose validating persona outcomes with insights from other sources, such as user interviews or market research, to ensure the validity of the final personas.

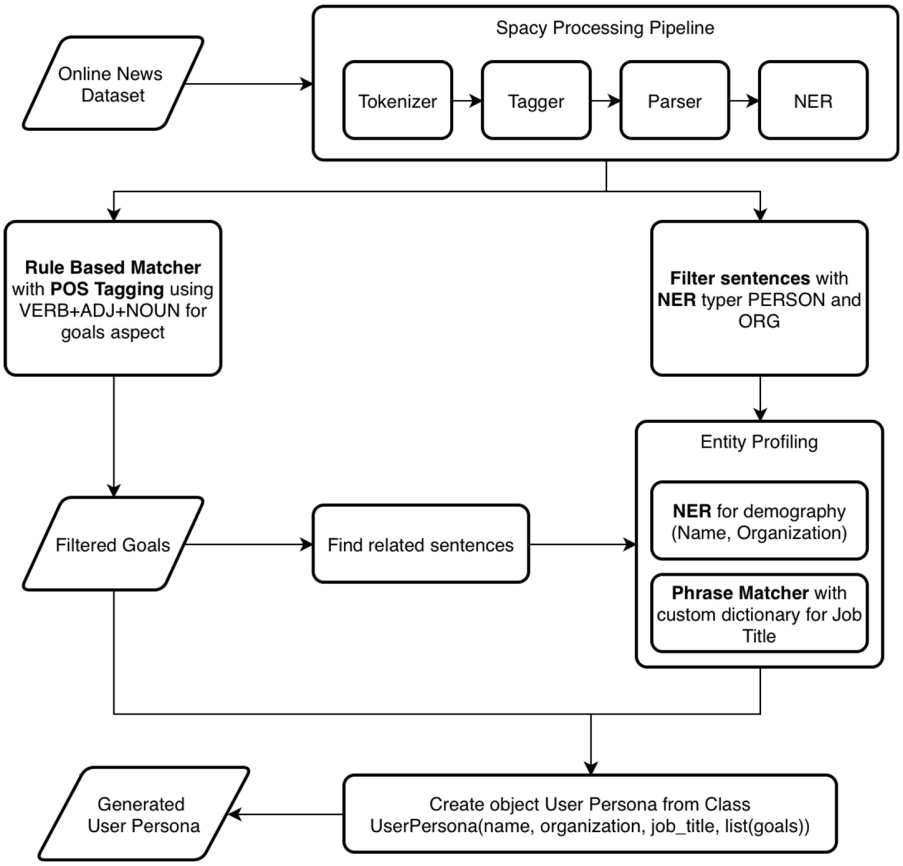


Fig. 1 Proposed Method

A. Formulation of user persona

In this phase, the components that comprise our user persona output are identified. Previous research has offered valuable insights into user personas. Cooper [32] emphasized the importance of identity and goals, whereas Mesgari et al. [26] outlined potential persona attributes, including demography (e.g., age, job, and education), psychographics (e.g., lifestyle, goals, needs, and intention), and interaction (e.g., clickstream and system interaction). Park and Kang [18]. Additionally, Olsen [33] provided a comprehensive guide detailing the aspects that should be included in user personas.

Furthermore, from previous research, such as [19], [27], and [32], we observed a commonality between them, with the most commonly used aspects being name and goals. Jobs and workplaces are also used as demographic information. Therefore, we added jobs and workplaces as additional information to enrich the persona. Drawing from these prior studies, we focused on extracting names, jobs, workplaces, and goals as the primary components of our user persona output. This component is the most consistently available in online news and directly corresponds to key aspects of the persona, namely identity, affiliation, and goals. This study emphasizes the feasibility of identifying core entities as proof of concept.

B. Data Collection

To test our proposed method, our dataset should include the target system, allowing us to replicate the software development process and validate the user persona generated by our proposed method. This aligns with prior research [34], which demonstrated the real-world application of developing a system and leveraging user personas to enhance user awareness and refine design, functionality, and user experience. The chosen target system, a city dashboard, employs user personas as a reference and guide to empathize with users, thereby enhancing design, functionality, and user experience. Therefore, we selected five topics based on mobile application trends from 2022 to 2023 [35]. These five topics are symptom checker, chatbot, smart hub, food delivery, and marketplace. We used these topics to find online news and imagined them as software development projects. In other words, our method was evaluated across five distinct projects: (1) symptom checker application, (2) chatbot application, (3) smart hub application, (4) food delivery application, and (5) marketplace application.

The system requirements for each project were not specified because we focused on evaluating the user persona created. However, since we identified the topics that define the application as a mobile application, we assume and anticipate that the project will be in development. We were more concerned about whether the user persona generated by our proposed method could capture valuable user personas in software development, as assessed by experts in the evaluation.

For our dataset, we sourced data from Google News, which automates global news aggregation every 15 minutes, offering diverse viewpoints from 4,500 sources without editorial bias [36], focusing on topics such as symptom checkers, chatbots, smart hubs, food delivery, and marketplaces, as outlined in [37]. We collected four news articles for each topic, resulting in 20 articles. To ensure relevance, we established specific criteria: the articles had to be in English, identified using keywords like "[topic] + 'app'" (e.g., "chatbot app"), include at least one person's name, and relate to at least one mobile application. In addition, we considered only articles containing more than 250 words.

C. Spacy Processing Pipeline

At this stage, we use SpaCy as our primary NLP tool. These include tokenization, lemmatization, part-of-speech (POS) tagging, dependency parsing, and NER. POS tagging and dependency parsing enable the identification of goal-related phrases by detecting grammatical structures. Meanwhile, NER is used to extract entity types relevant to the construction of personas. Consequently, every text processed by SpaCy passes through this pipeline, where all linguistic attributes are embedded and stored in a doc format.

D. Rule-Based Matcher for Identifying Goal Aspects

After obtaining Universal POS tag attributes using SpaCy, we applied a Rule-Based Matcher to detect goal-related phrases following the linguistic pattern VERB + ADJECTIVE + NOUN, as adopted from [23]. The total number of goals per article varies depending on the occurrences within the text. These filtered goals represent goal-oriented phrases obtained after POS tagging, which describe the intended actions or desired outcomes of stakeholders mentioned in the news text.

This process is illustrated with the following example. Consider the following sentence: *"According to Lejbolle, using a patient portal and EHR application from the same vendor provides more convenience and security."* In this case, the phrase "*provides more convenience*" is identified as a goal-related expression because it matches the defined linguistic pattern of VERB + ADJECTIVE + NOUN within our SpaCy rule-based matcher.

E. NER to Identify Name Aspects

We then continue to search for names available in the news using NER. The proposed NER process considers two entity types: PERSON and ORG. This dual-entity approach accommodates instances where names can be associated with individuals (PERSON) or organizations (ORG), which may have relevant goals articulated within the news articles.

For illustration, consider the following sentence: *"According to Lejbolle, using a patient portal and EHR application from the same vendor provides more convenience and security.*" In this example, the NER process identifies "*EHR*" and *"Lejbolle"* as entities, corresponding to an organization and a person, respectively. The extracted entities are then collected as potential candidates for constructing the user persona.

F. Entity Profiling

Our data collection captures the marked items and retrieves the sentences from which these items are sourced. These sentences are crucial for confirming or identifying any remaining entities or goals. Additionally, these sentences were leveraged to establish relationships and connect the extracted names, jobs, workplaces, and goals. If multiple entities (e.g., a person, a job, and a goal) appear in the same sentence, they are assumed to be related. However, this does not

ensure that the extracted entities are correctly linked. This study does not fully automate the linking of entities into personas. The construction of complete user personas still requires human involvement.

Now that we have acquired names and goals, our focus shifts to jobs. We use a custom dictionary with SpaCy's phrase matcher to identify job titles. Our custom dictionary comprises 450 job titles sourced from [38]. This enables us to efficiently identify job-related mentions within news articles.

For workplaces, NER is used. By implementing these methods, we ensure that every aspect of our user personas is comprehensively addressed at this stage. To facilitate data organization, we have designed a user persona class to store this information, which is subsequently aggregated to create a JSON representation.

G. Evaluation

For the evaluation, we have enlisted the expertise of five qualified individuals who meet specific criteria: (1) each expert possesses experience in creating or understanding user personas, and (2) proficiency in English is a prerequisite. We provide details regarding their expertise, years of experience, and background training to demonstrate the sufficiency and qualification of these experts. Each expert selected has a background in UI/UX design or project management, with experience using user personas within their respective roles. Despite the small number of experts , their diverse backgrounds and expertise in persona development ensure that the validation procedure is complete and reliable.

The evaluation process involves several steps. Initially, two distinct professionals manually check each dataset, ensuring a comprehensive evaluation of each news article. Subsequently, the experts identify user persona aspects, such as names, jobs, workplaces, and goals. Based on their findings, a ground truth is created to accurately represent the identified aspects. Subsequently, a meeting is held with the specialists to verify and enhance this ground truth. The ground truth serves as a benchmark for actual user personas, whereas the JSON results generated by the proposed method represent the predicted user personas. By comparing the two, a confusion matrix is generated [27] Finally, we solicit expert feedback on the interaction design of our system and evaluate the method. This feedback helps assess the process's usability and usefulness, informing potential improvements in both aspects. This comprehensive evaluation approach ensures a thorough assessment of the performance and practical usability of our method.

The confusion matrix will have TP, FP, TN, and FN values. In our case, TP represents the number of ground truths correctly identified by the proposed method, and FP represents the number of aspects incorrectly identified by the proposed method. However, in the ground truth, FN refers to the number of ground truths that are not identified using the proposed method. We did not use TN because no exclusion or negative target exists in the ground truth or the proposed system. This matrix enables the measurement of performance metrics, including recall, precision, and F1 score. These metrics provide valuable insights into the effectiveness of the proposed method. The precision, recall, and F1-score formulas are as follows:

$$recall = \frac{\mathrm{Tp}}{\mathrm{Tp+Fn}} \quad (1)$$

$$precision = \frac{\mathrm{Tp}}{\mathrm{Tp+Fp}} \quad (2)$$

$$f1 - score = \frac{2\mathrm{Tp}}{2\mathrm{Tp+Fp+Fn}} \quad (3)$$

III. RESULTS

The results are presented in subsections, encompassing outcomes from various stages of our study, including data collection, SpaCy Processing Pipeline, rule-based matcher, NER for name identification, entity profiling, and evaluation.

A. Data Collection

We successfully retrieved 20 news articles spanning five distinct topics from Google News. To streamline references throughout the study, we assigned codes to each topic: SC, CB, SH, FD, and MP. Table 1 presents a comprehensive overview of our data collection results.

TABLE 1
DATASET IDENTIFICATION

No	Topics	Title	Code	Sentence count	Word count
1	Symptom Checker	Columbia Memorial Health launches the region's first digital "symptom checker" powered by Mediktor	SC1	34	602
2		Study finds big differences between top symptom checker apps	SC2	18	471
3		Japanese startup Ubie brings AI symptom checker to the US	SC3	17	351
4		I used Alexa's symptom checker to track my health for a month — here are the pros and cons	SC4	46	1355
5	Chatbot	Infobip creates AI-powered chatbot for Uber, enabling its Delhi-NCR customers to book rides via WhatsApp	CB1	34	873
6		'Chat' with Musk, Trump or Xi: Ex-Googlers want to give the public AI	CB2	89	1491
7		Nomura Securities' "OneStock" App adopted Allganize's AI Chatbot "Alli"	CB3	31	639
8		'PetTalk' ChatBot in South Korea Aims to Help You 'Talk' to Your Dog	CB4	22	505
9	Smart hub	IKEA introduces DIRIGERA and the new IKEA Home smart app	SH1	23	536
10		Exclusive: Ikea's new Dirigera hub and Home smart app deliver big improvements	SH2	107	2323
11		Ikea overhauls its smart home with new Matter-ready hub and app	SH3	98	2030
12		Matter: what you need to know	SH4	152	2633
13	Food Delivery	Food delivery service Deliveroo has arrived in Qatar	FD1	27	633
14		Easypaisa to Launch Food Delivery Apps in Collaboration with Blink	FD2	16	377
15		DoorDash Aims to Outcompete Rivals With SMB Perks Program	FD3	23	563
16		Joanna's Take: The future of food delivery apps hangs in the balance	FD4	27	806
17	Market Place	PrettyLittleThing Marketplace app: Greenwashing or a step forward?	MP1	24	551
18		Stripe launches App Marketplace, scripts and tools incorporating third-party SaaS apps that work alongside Stripe	MP2	22	704
19		Stripe's new apps marketplace brings third-party tools directly into Stripe	MP3	16	360
20		Aviation Marketplace FLYJETS Debuts Mobile App and New Scheduling, FLYCalendar and FLYGreen Features	MP4	31	686

B. Rule-Based Matcher

The SpaCy processing pipeline yields data in Python's "doc" format, encompassing various attributes, including lemmatization, POS tags, and dependency tags. A rule-based matcher process is performed after the SpaCy processing pipeline. The rule-based matcher enabled us to extract goals and sentences containing them. An illustrative example of the results is provided in Table 2. The goals serve as the output, while the input sentences containing these goal phrases are preserved. Each sentence contains one representative goal phrase. The proposed method successfully retrieved 311 goals across all 20 news articles.

TABLE 2
RULE-BASED MATCHER RESULT

No	Input	Output
1	Amazon and Google have said their existing smart speakers and displays will be upgraded this way, and we should hear more about other companies plans now that the final spec is out.	existing smart speakers
2	Ecosystems will want new ways to show the benefits of their user experiences.	want new ways
3	Eves smart plugs and Yales locks are two examples Yale using swappable modules for its locks and Eve adding Thread radios to many of its devices.	using swappable modules
4	Finally, Matter-enabled smartphone apps such as Apple Home, Google Home, Amazon Alexa, or Samsung SmartThings are also controllers.	enabled smartphone apps
5	Like a Thread border router, controller functionality can be built into many devices, and you can have multiple controllers.	have multiple controllers

C. NER to Identify the Name Aspect

This section focuses on identifying names mentioned in news articles. Our NER process encompasses two entity types: PERSON and ORG. Using this method, we successfully retrieved 246names. Notably, specific names may appear in multiple articles within the same topic.

Notably, our approach accommodates both NER types, PERSON and ORG, as some of our experts argued for their inclusion. However, during the evaluation stage, we will delve into the disadvantages of this dual-entity approach, providing a comprehensive analysis of its impact on performance.

D. Entity Profiling

This section explores entity profiling, which is a crucial step in our methodology. Following the acquisition of names and goals, the process extends to identifying the jobs and workplaces associated with each entity mentioned in the names. This endeavor resulted in the retrieval of 41 job titles and 72 workplaces. To perform this analysis, we used sentences containing both goals and names.

The outcomes of this entity profiling stage were integrated into a user persona class. As a result, we successfully generated user personas in four distinct types, as shown in Fig. 2. Fig. 2a shows user personas that solely include names. Fig. 2b shows user personas that include information about names and workplaces. Fig. 2c shows the user personas, which include names and job titles. Fig. 2d shows the user personas, including names, job titles, and workplaces.

(a)

```
{'name': 'Michele Turner',
 'work': '',
 'goals': [],
 'job_title': ''}
```

(b)

```
{'name': 'Michelle Mindala-Freeman',
 'work': 'CSA',
 'goals': [],
 'job_title': ''}
```

(c)

```
{'name': 'Yoshinori Abe',
 'work': [],
 'goals': [],
 'job_title': 'founder'}
```

(d)

```
{'name': 'Rebecca Töreman',
 'work': 'Ikea Home Smart',
 'goals': [],
 'job_title': 'business
leader'}
```

(e)


```
[
  {"name": "Kevin Kraus", "work": "Yale", "goals": [], "job_title": []},
  {"name": "Kevin Po", "work": "Google Nest", "goals": [], "job_title": []},
  {"name": "Klein", "work": [], "goals": [], "job_title": []},
  {"name": "Matter", "work": "Unify SDK", "goals": ["existing smart home"], "job_title": "controller"},
  {"name": "Matter 1.0", "work": [], "goals": [], "job_title": []},
  {"name": "Matter Controller", "work": "Samsung SmartThings", "goals": [], "job_title": []},
  {"name": "Matter Controllers", "work": [], "goals": [], "job_title": []},
  {"name": "Matter SoC", "work": "Infineon", "goals": [], "job_title": []},
  {"name": "Michele Turner", "work": [], "goals": [], "job_title": []},
  {"name": "Michelle Mindala-Freeman", "work": "CSA", "goals": [], "job_title": []},
  {"name": "Nest", "work": [], "goals": [], "job_title": []},
  {"name": "Nest Hub", "work": [], "goals": [], "job_title": []},
  {"name": "Po", "work": [], "goals": ["provides common controls"], "job_title": []}
]
```


Fig. 2 Types of User Persona Produced (a) user persona with only name as identified aspect, (b) user persona with name and workplace identified, (c) user persona with name and job title identified, and (d) user persona with name, workplace, and job title identified. (e) the example structured user personas generates by the system. This relationship will serve as the basis for the user persona candidate in Figure 3.

As shown in Fig. 2e, The profiling process assigns attributes such as name**,** job**,** and organization to each identified entity, forming the user persona's basic structure. However, the output may vary depending on the completeness of the detected information. Notably, while some of these personas include goals, others may not. Additionally, we introduced an entity called "user," which serves as a default placeholder for storing goals when no identifiable or named entities are present in the sentences. In cases where multiple entities appear within the same sentence or article, each detected name is treated as an independent entity. When indications are insufficient to determine relationships, the final connection is manually reviewed by humans during the validation process.

E. Evaluation

The evaluation was conducted using two complementary approaches. Precision, recall, and F1-score were measured to assess the accuracy of entity and goal identification. We also performed an expert-based evaluation to examine whether the personas generated could support the requirements elicitation process.

In this section, we present the evaluation results of our proposed method using performance metrics such as recall, precision, and F1-score, which are calculated by comparing the system-generated results with the expert annotations that serve as the ground truth. The evaluation is conducted on two levels: (i) Article-Level Evaluation*,* which assesses the overall performance of the system for each news article, focusing on the accuracy of persona generation at the document level, as shown in Table 3, and (ii) User Persona Aspect-Level Evaluation, as presented in Table 4, which measure the performance of the system in identifying each individual attribute of a user persona, namely: name, job, organization, and goal. We applied a macro-averaging approach, where the mean value of all aspects was used to represent the overall system performance. This method ensures that each aspect (name, job, organization, and goal) contributes equally to the final evaluation, regardless of its frequency in the dataset.

TABLE 3
EVALUATION RESULT IN GENERAL

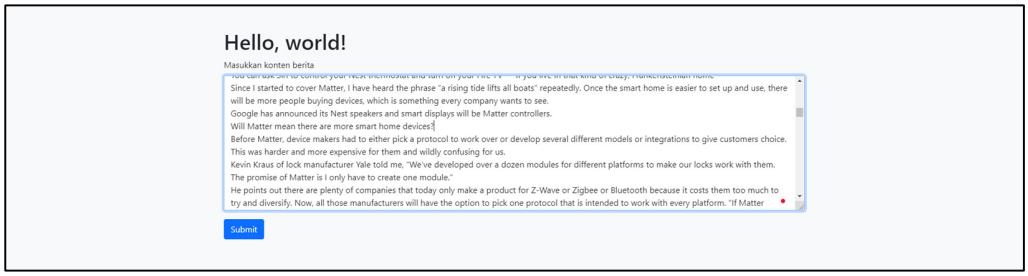
Code	Identified by experts				Identified by the proposed method				performance		
	name	job	org	goals	name	job	org	goals	recall	precision	f1
CB1	5	4	2	14	13	5	5	14	0.720	0.486	0.581
CB2	7	6	2	13	24	3	5	26	0.393	0.190	0.256
CB3	3	2	2	4	10	2	4	4	0.909	0.500	0.645
CB4	1	1	1	8	6	3	2	7	0.727	0.444	0.552
FD1	2	2	1	13	7	2	5	17	0.889	0.516	0.653
FD2	2	2	2	5	7	1	2	7	0.727	0.471	0.571
FD3	3	3	5	2	11	2	2	11	0.692	0.346	0.462
FD4	1	1	3	5	11	1	4	11	0.800	0.296	0.432
MP1	5	4	2	8	8	1	3	11	0.684	0.565	0.619
MP2	1	1	1	9	7	2	2	10	0.750	0.429	0.545
MP3	1	1	1	7	6	2	2	7	0.700	0.412	0.519
MP4	1	2	1	7	11	3	2	14	0.727	0.267	0.390
SC1	4	4	2	5	9	4	3	7	0.667	0.435	0.526
SC2	8	1	1	9	8	1	2	13	0.579	0.458	0.512
SC3	2	2	3	7	9	1	1	10	0.571	0.381	0.457
SC4	3	3	6	20	15	2	4	29	0.656	0.420	0.512
SH1	1	1	2	12	6	1	2	19	0.750	0.429	0.545
SH2	2	2	6	18	23	2	4	30	0.714	0.339	0.460
SH3	2	2	9	12	18	1	7	23	0.640	0.327	0.432
SH4	6	1	9	27	37	2	11	41	0.698	0.330	0.448
Total	60	45	61	205	246	41	72	311			
Average									0.700	0.402	0.506

TABLE 4
PERFORMANCE MEASURE FOR EACH USER PERSONA ASPECT

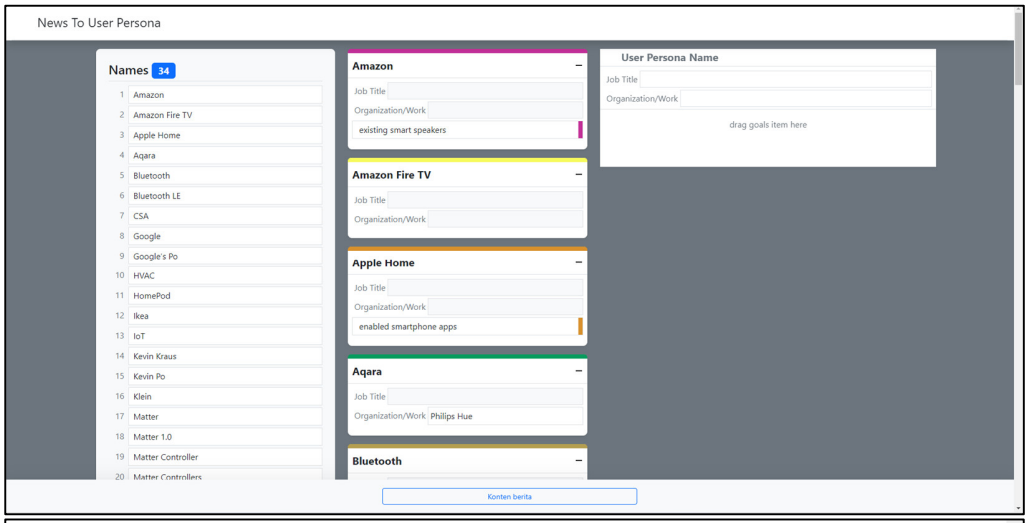
Code	Name			Job			Workplace			Goals		
	recall	precision	F1	recall	precision	F1	recall	precision	F1	recall	precision	F1
CB1	0.600	0.231	0.333	0.500	0.400	0.444	0.400	0.400	0.400	0.786	0.786	0.786
CB2	0.714	0.208	0.323	0.167	0.333	0.222	0.100	0.200	0.133	0.308	0.154	0.205
CB3	1.000	0.300	0.462	1.000	1.000	1.000	0.667	0.500	0.571	0.750	0.750	0.750
CB4	1.000	0.167	0.286	1.000	0.333	0.500	0.000	0.000	0.000	0.750	0.857	0.800
FD1	1.000	0.286	0.444	1.000	1.000	1.000	0.333	0.200	0.250	0.846	0.647	0.733
FD2	1.000	0.286	0.444	0.500	1.000	0.667	0.500	0.500	0.500	0.800	0.571	0.667
FD3	1.000	0.273	0.429	0.667	1.000	0.800	1.000	1.000	1.000	1.000	0.182	0.308
FD4	1.000	0.091	0.167	1.000	1.000	1.000	1.000	0.250	0.400	1.000	0.455	0.625
MP1	1.000	0.625	0.769	0.250	1.000	0.400	0.333	0.333	0.333	0.750	0.545	0.632
MP2	1.000	0.143	0.250	1.000	0.500	0.667	0.250	0.500	0.333	0.667	0.600	0.632
MP3	1.000	0.167	0.286	1.000	0.500	0.667	0.000	0.000	0.000	0.714	0.714	0.714
MP4	1.000	0.091	0.167	1.000	0.667	0.800	0.000	0.000	0.000	0.714	0.357	0.476
SC1	0.750	0.333	0.462	0.750	0.750	0.750	0.333	0.333	0.333	0.600	0.429	0.500
SC2	0.500	0.500	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.778	0.538	0.636
SC3	1.000	0.222	0.364	0.500	1.000	0.667	0.000	0.000	0.000	0.714	0.500	0.588
SC4	1.000	0.200	0.333	0.333	0.500	0.400	0.286	0.500	0.364	0.750	0.517	0.612
SH1	1.000	0.167	0.286	0.000	0.000	0.000	0.333	0.500	0.400	0.833	0.526	0.645
SH2	0.500	0.043	0.080	0.500	0.500	0.500	0.500	0.500	0.500	0.889	0.533	0.667
SH3	1.000	0.111	0.200	0.500	1.000	0.667	0.667	0.286	0.400	0.917	0.478	0.629
SH4	0.833	0.135	0.233	0.000	0.000	0.000	0.400	0.364	0.381	0.778	0.512	0.618
Average	0.895	0.229	0.341	0.583	0.624	0.558	0.355	0.318	0.315	0.767	0.533	0.611

Figure 3 illustrates the system demonstration of the proposed method. A web-based interface was developed by integrating the proposed method into a Django-based framework to facilitate the evaluation and presentation of results. The interface follows an input–process–output concept, where users can input and submit online news text for processing. The extraction results are then displayed in three columns: (a) identified aspects, (b) user persona candidates derived from relational analysis between entities, and (c) final user persona composition. Users can refine and adjust the personas generated by the proposed method through this interface. This interactive mechanism also addresses the phenomenon of goal aspects that may appear without a corresponding entity, enabling users to manually establish appropriate relationships between aspects and thereby improving the overall evaluation quality and accuracy of persona construction.

(a)



(b)



(c)

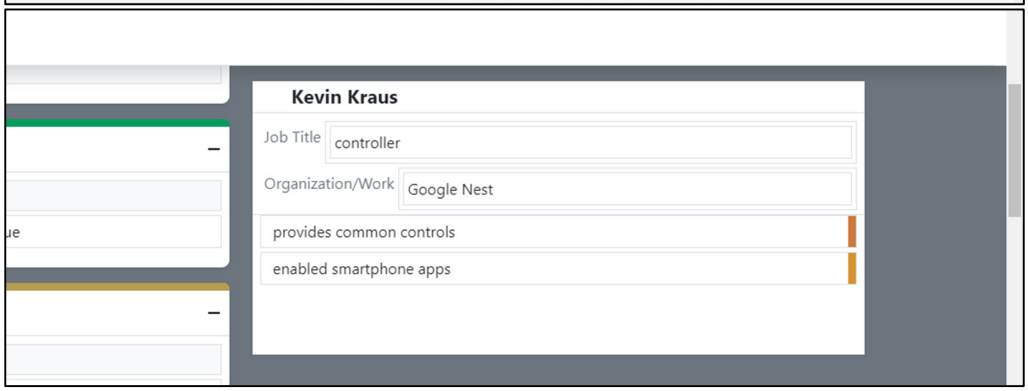


Fig. 3 System demonstration of our method: (a) user can input the online news text and submit it, (b) the extraction result is separated into three columns: the first column shows the identified aspects, the second column is user persona candidates, and the third one is where user can drag and drop necessary aspects to create a persona, (c) user persona result.

We evaluate the performance of the proposed method for each user persona aspect, including names, jobs, workplaces, and goals. This approach provides insights into the general and specific performance of the proposed

method. Table 3 summarizes the general performance metrics, indicating an average recall of 0.70, a precision of 0.40, and an F1-score of 0.50. Figure 3 provides further insights into the performance, which illustrates the average performance for each topic. Across all topics, the recall value consistently surpasses precision.

This study aimed to assess the efficacy of the proposed method in facilitating the elicitation of requirements. Most experts strongly agreed that our method aids in eliciting requirements and found it to be highly beneficial. One expert (20%) strongly agreed that the system was easy to use, while three experts (60%) expressed agreement, and one expert remained neutral in their assessment. Furthermore, when assessing the user-friendliness of the system, four experts (80%) found it easy to use, and one expert (20%) found it very easy. Regarding the quality, four experts (80%) deemed them reasonable, and one expert expressed neutrality. Furthermore, when evaluating whether the proposed method can facilitate the requirements elicitation process, four experts (80%) strongly agreed, and one expert (20%) expressed neutrality. Regarding the level of assistance provided by the proposed method, all experts (100%) found it helpful.

The expert feedback on the usability of our interface generally conveyed positive opinions regarding its ease of use. This finding is consistent with [39] the significance of user-centered design in achieving high usability ratings [39]. However, one expert's neutral perspective suggested potential areas for interface enhancements to better align with user expectations. Nevertheless, a minority of experts expressed a neutral stance, emphasizing the importance of incorporating aspects related to challenges or user avoidance behaviors. This study addressed whether the proposed method effectively aids the requirements elicitation process. Most experts strongly agreed that the proposed method contributes to requirements elicitation and found it highly helpful.

IV. Discussion

The following section discusses the study's findings, limitations, and comparisons to and links with related earlier studies.

A. The Finding

This study demonstrates that online news articles can serve as a complementary source for eliciting requirements, although selecting relevant articles presents a considerable challenge. Not all news articles contain names, job titles, workplaces, or career goals. This requires the screening of news articles to ensure relevance. Based on the performance evaluation, the proposed method demonstrated that the recall value was consistently higher (0.700) than the precision value (0.402), with an overall F1 score of 0.506. These results indicate that the proposed method effectively captures a broad set of user persona attributes; however, it often yields predictions that do not align with expert annotations. This trade-off highlights the common challenge of extracting needs from unstructured texts, where increased coverage often comes at the expense of accuracy [27], [40], [41]. The generated personas serve as preliminary stakeholder-based profiles that provide developers with initial insights into potential users and their goals.

Further analysis of each persona aspect shows different performance patterns. The name aspect achieved very high recall (0.895) but low accuracy (0.229), primarily due to the system's tendency to identify both the individual's name and the organization's name. At the same time, the expert mostly only annotated the individual. This result indicates that the proposed method detected most of the names identified by the experts but produced too many predictions. The job aspect demonstrates moderate performance (recall 0.582, precision 0.624), which is limited by a static dictionary of 450 job titles that does not include new professions such as "influencer" or "business leader." Overall, the workplace aspect was the weakest (recall 0.355, precision 0.318), reflecting the difficulty in recognizing the organization's name in news articles. Finally, the goal aspect showed a higher recall (0.767) than precision (0.553), a pattern similar to that observed in the name aspect. These findings emphasize the need for more adaptive methods to balance scope and accuracy across different dimensions of personas.

The variation in precision, recall, and F1-score across attributes in Table 4 can be explained by distributing entities in online news articles. Names, jobs, and organizations often appear explicitly and consistently. High or even perfect scores in some cases. Conversely, goals are more implicit and have diverse syntactic structures, affecting detection accuracy. Because this framework employs a rule-based NLP approach rather than a trained model, perfect scores do not necessarily indicate overfitting; instead, they reflect the effectiveness of the rules in specific sentence patterns.

In comparison to previous Facebook or Twitter. In contrast, this study uses online news articles as a new data source for eliciting requirements research, this study makes several important contributions. Studies [24] [27] and [39] rely on social media profiles to build personas, using predefined dictionaries or user quotes from platforms such as and introducing goal extraction through a point-of-sale (POS)-tagging pattern [42], [43], [44]. Unlike previous studies, this study explicitly evaluated performance using precision, recall, and F1 score metrics, providing transparent and measurable insights into the effectiveness of the system. However, no direct baseline comparison was conducted because of differences in data sources across existing studies, such as app reviews or social media posts. The findings provide a methodological foundation for future comparative evaluations using standardized evaluation metrics.

The findings of this study have practical implications for software engineering, particularly in the early stages of requirements elicitation. Online news articles offer cost-effective and easily accessible resources for generating initial user personas. UX developers and designers can gain insights into users' context, expectations, and goals by using social narratives reflected in the media. The framework developed in this study demonstrates how the method can be operationalized in practice. The feedback from experts shows that the system is generally easy to use and capable of supporting requirements elicitation activities, although there are some suggestions for improvement.

B. Limitations of the Study

This study has some limitations. First, the dataset is relatively small (comprising 20 articles), limiting the generalizability of the findings. Relying on a single source, such as Google News, can lead to bias. It can distort the representation of personas, as topicality and editorial selection influence news coverage. Second, the performance in the work and workplace aspects is limited by the limitations of the NER model and the static job dictionary. This approach has only been tested in the context of mobile apps; thus, its applicability to other domains remains uncertain.

This study individually evaluates each persona aspect using precision, recall, and F1-score. However, input may simultaneously contain multiple aspects (e.g., name, job, workplace, and goal), and the Hamming Loss metric could be employed to assess overall prediction errors across all aspects [45]. Future work will consider incorporating this metric to enhance the accuracy of the assessment of persona extraction.

Future research should aim to overcome these limitations by using larger and more diverse datasets. Methodological improvements could include integrating transformer-based models, such as BERT, to capture richer contextual information, particularly on aspects of work and objectives. Beyond methodological improvements, extending this approach to other domains (e.g., healthcare, fintech) will be important for assessing scalability and generalization. In the long run, integrating news-based persona creation into the automated flow of requirements elicitation can significantly reduce the reliance on manual methods, driving a vision of more efficient and data-driven software development. This framework should be extended in future research by incorporating additional NER types and contextual features to capture broader persona dimensions, including demographic, behavioral, and psychographic aspects.

V. CONCLUSIONS

This study offers a new approach to generating user personas from online news data by extracting key aspects of user personas, such as names, jobs, workplaces, and goals, using natural language processing techniques. Named entity recognition (NER) is used to identify named entities, such as names and organizations. Aspects of jobs are obtained through a custom dictionary with SpaCy's Phrase Matcher, while goals are extracted using a rule-based pattern of VERB + ADJECTIVE + NOUN. The evaluation results and usability feedback demonstrate the proposed system's practicality as a complementary tool for understanding user needs.

Although the study's findings are promising, there is still room for improvement. Further research should focus on improving the precision of entity recognition, automating the identification of job aspects beyond static dictionaries, and developing more sophisticated mechanisms for goal extraction. Expanding to specific domains within mobile apps will be a crucial step in assessing generalizations and refining their implementation.

Author Contributions: *Halim Wildan Awalurahman*: Software, Investigation, Data Curation, Writing - Original Draft Preparation. *Indra Kharisma Raharjana*: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing - Original Draft Preparation, Supervision, Funding Acquisition. *Kartono Kartono*: Writing - Review and Editing, Supervision. *Shukor Sanim Mohd Fauzi*: Methodology, Validation, Writing - Review and Editing.

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

Funding: This work was supported by Direktorat Riset, Teknologi, dan Pengabdian Kepada Masyarakat Kementrian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia through Penelitian Fundamental Reguler (PFR) under Grant 0536/E5/PG.02.00/2023.

Conflicts of Interest: Awalurahman, Raharjana, and Fauzi are members of the Editorial Team, 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 datasets generated and analyzed during this study are publicly available at the following

repository: https://github.com/hawalurahman/news-to-user-persona

Informed Consent: Not applicable.

Institutional Review Board Statement: Not applicable.

Animal Subjects: There were no animal subjects.

ORCID:
Halim Wildan Awalurahman: https://orcid.org/0009-0005-8136-9970
Indra Kharisma Raharjana: https://orcid.org/0000-0002-0622-3374
Kartono Kartono: https://orcid.org/0009-0005-8480-7710
Shukor Sanim Mohd Fauzi: https://orcid.org/0000-0003-4333-7853

REFERENCES

[1] F. M. Khan, J. A. Khan, M. Assam, A. S. Almasoud, A. Abdelmaboud, and M. A. M. Hamza, "A Comparative Systematic Analysis of Stakeholder's Identification Methods in Requirements Elicitation," *IEEE Access*, vol. 10, pp. 30982–31011, 2022, doi: 10.1109/ACCESS.2022.3152073.

[2] A. Ahmad *et al.*, "A Systematic Literature Review on Using Machine Learning Algorithms for Software Requirements Identification on Stack Overflow," *Security and Communication Networks*, vol. 2020, pp. 1–19, Jul. 2020, doi: 10.1155/2020/8830683.

[3] F. N. J. Muhamad, S. H. Ab Hamid, H. Subramaniam, R. Abdul Rashid, and F. Fahmi, "Fault-Prone Software Requirements Specification Detection Using Ensemble Learning for Edge/Cloud Applications," *Applied Sciences (Switzerland)*, vol. 13, no. 14, 2023, doi: 10.3390/app13148368.

[4] F. Anvari, H. M. T. Tran, D. Richards, and M. Hitchens, "Towards a method for creating personas with knowledge and cognitive process for user centered design of a learning application," *Proceedings - 2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering, CHASE 2019*, no. 1, pp. 123–130, 2019, doi: 10.1109/CHASE.2019.00037.

[5] A. Aldave, J. M. Vara, D. Granada, and E. Marcos, "Leveraging creativity in requirements elicitation within agile software development: A systematic literature review," *Journal of Systems and Software*, vol. 157, 2019, doi: 10.1016/j.jss.2019.110396.

[6] I. Puspitasari, N. Nuzulita, and C.-S. Hsiao, "Agile User-Centered Design Framework to Support the Development of E-Health for Patient Education," in *Computer and Information Science and Engineering: Volume 16*, R. Lee, Ed., Cham: Springer Nature Switzerland, 2024, pp. 131–144. doi: 10.1007/978-3-031-57037-7_10.

[7] P. Kamthan, "Using Personas to Support the Goals in User Stories," in *2015 12th International Conference on Information Technology - New Generations*, IEEE, Apr. 2015, pp. 770–770. doi: 10.1109/ITNG.2015.136.

[8] B. J. Jansen, S. G. Jung, L. Nielsen, K. W. Guan, and J. Salminen, "How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment," *Pacific Asia Journal of the Association for Information Systems*, vol. 14, no. 3, pp. 1–28, 2022, doi: 10.17705/1pais.14301.

[9] T. Huynh, A. Madsen, S. McKagan, and E. Sayre, "Building personas from phenomenography: a method for user-centered design in education," *Information and Learning Sciences*, vol. 122, no. 11–12, pp. 689–708, Jul. 2021, doi: 10.1108/ILS-12-2020-0256.

[10] M. Mesgari, C. Okoli, and A. O. De Guinea, "Creating Rich and Representative Personas by Discovering Affordances," *IEEE Transactions on Software Engineering*, vol. 45, no. 10, pp. 967–983, 2019, doi: 10.1109/TSE.2018.2826537.

[11] J. Choma, L. A. M. Zaina, and D. Beraldo, "UserX story: Incorporating UX aspects into user stories elaboration," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9731, pp. 131–140, 2016, doi: 10.1007/978-3-319-39510-4_13.

[12] A. Cooper, *Design for Pleasure*. 2004.

[13] A. Hess, P. Diebold, and N. Seyff, "Understanding information needs of agile teams to improve requirements communication," *J Ind Inf Integr*, vol. 14, no. November 2017, pp. 3–15, 2019, doi: 10.1016/j.jii.2018.04.002.

[14] J. Salminen, K. Wenyun Guan, S. G. Jung, and B. Jansen, "Use Cases for Design Personas: A Systematic Review and New Frontiers," *Conference on Human Factors in Computing Systems - Proceedings*, 2022, doi: 10.1145/3491102.3517589.

[15] J. Salminen, S. G. Jung, and B. Jansen, "Developing Persona Analytics Towards Persona Science," *International Conference on Intelligent User Interfaces, Proceedings IUI*, pp. 323–344, 2022, doi: 10.1145/3490099.3511144.

[16] P. Losana, J. W. Castro, X. Ferre, E. Villalba-Mora, and S. T. Acuña, "A Systematic Mapping Study on Integration Proposals of the Personas Technique in Agile Methodologies," *Sensors*, vol. 21, no. 18, p. 6298, Sep. 2021, doi: 10.3390/s21186298.

[17] N. T. Khanh, J. Daengdej, and H. H. Arifin, "Human stories - A new written technique in Agile Software requirements," *ACM International Conference Proceeding Series*, pp. 15–22, 2017, doi: 10.1145/3056662.3056680.

[18] D. Park and J. Kang, "Constructing Data-Driven Personas through an Analysis of Mobile Application Store Data," *Applied Sciences (Switzerland)*, vol. 12, no. 6, 2022, doi: 10.3390/app12062869.

[19] J. McGinn and N. Kotamraju, "Data-driven persona development," *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1521–1524, 2008, doi: 10.1145/1357054.1357292.

[20] A. Hinderks, F. J. Domínguez Mayo, J. Thomaschewski, and M. J. Escalona, "Approaches to manage the user experience process in Agile software development: A systematic literature review," *Inf Softw Technol*, vol. 150, no. October 2020, p. 106957, 2022, doi: 10.1016/j.infsof.2022.106957.

[21] M. R. Dewi, I. K. 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*, New York, NY, USA: ACM, Feb. 2021, pp. 81–87. doi: 10.1145/3457784.3457796.

[22] A. C. Emcha, Widyawan, and T. B. Adji, "Quotation extraction from Indonesian online news," *2019 International Conference on Information and Communications Technology, ICOIACT 2019*, pp. 408–412, 2019, doi: 10.1109/ICOIACT46704.2019.8938558.

[23] I. K. Raharjana, D. Siahaan, and C. Fatichah, "User Story Extraction from Online News for Software Requirements Elicitation: A Conceptual Model," *JCSSE 2019 - 16th International Joint Conference on Computer Science and Software Engineering: Knowledge Evolution Towards Singularity of Man-Machine Intelligence*, pp. 342–347, Jul. 2019, doi: 10.1109/JCSSE.2019.8864199.

[24] M. R. Dewi, I. K. 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*, New York, NY, USA: ACM, Feb. 2021, pp. 81–87. doi: 10.1145/3457784.3457796.

[25] A. C. Emcha, Widyawan, and T. B. Adji, "Quotation extraction from Indonesian online news," *2019 International Conference on Information and Communications Technology, ICOIACT 2019*, pp. 408–412, 2019, doi: 10.1109/ICOIACT46704.2019.8938558.

[26] M. Mesgari, C. Okoli, and A. O. De Guinea, "Creating Rich and Representative Personas by Discovering Affordances," *IEEE Transactions on Software Engineering*, vol. 45, no. 10, pp. 967–983, 2019, doi: 10.1109/TSE.2018.2826537.

[27] 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.

[28] E. Trisnawati, I. K. Raharjana, Taufik, A. H. Basori, N. A. Alghanmi, and A. B. F. Mansur, "Analyzing Variances in User Story Characteristics : A Comparative Study of Stakeholders with Diverse Domain and Technical Knowledge in Software Requirements Elicitation," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 1, pp. 110–125, 2024, doi: 10.20473/jisebi.10.1.110-125.

[29] Y. Wang *et al.*, "Who uses personas in requirements engineering: The practitioners' perspective," *Inf Softw Technol*, vol. 178, p. 107609, 2025, doi: https://doi.org/10.1016/j.infsof.2024.107609.

[30] B. J. Jansen, S.-G. Jung, J. Salminen, K. W. Guan, and L. Nielsen, "Strengths and Weaknesses of Persona Creation Methods: Guidelines and Opportunities for Digital Innovations," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021. doi: 10.24251/HICSS.2021.604.

[31] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, "spaCy: Industrial-strength Natural Language Processing in Python," 2020, *Zenodo*. doi: 10.5281/zenodo.10009823.

[32] A. Cooper, *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity*. Pearson Higher Education, 2004.

[33] G. Olsen, "Persona creation and usage toolkit." Accessed: Feb. 23, 2023. [Online]. Available: https://decampou.com/wp-content/uploads/2004/10/Guide-creation-Persona.pdf

[34] G. W. Young, R. Kitchin, and J. Naji, "Building City Dashboards for Different Types of Users," *Journal of Urban Technology*, vol. 28, no. 1–2, pp. 289–309, 2021, doi: 10.1080/10630732.2020.1759994.

[35] A. Jay, "16 Mobile App Trends for 2022/2023 and Beyond: Top Forecasts According to Experts - Financesonline.com."

[36] D. A. Weaver and B. Bimber, "Finding News Stories: A Comparison of Searches Using Lexisnexis and Google News," *Journal Mass Commun Q*, vol. 85, no. 3, pp. 515–530, Sep. 2008, doi: 10.1177/107769900808500303.

[37] A. Jay, "16 Mobile App Trends for 2022/2023 and Beyond: Top Forecasts According to Experts - Financesonline.com."

[38] T. Gerencer, "450 Job Titles for Professional Positions [Current & Desired]."

[39] S.-G. Jung, J. Salminen, and B. J. Jansen, "Giving Faces to Data: Creating Data-Driven Personas from Personified Big Data," in *Companion Proceedings of the 25th International Conference on Intelligent User Interfaces*, in IUI '20 Companion. New York, NY, USA: Association for Computing Machinery, 2020, pp. 132–133. doi: 10.1145/3379336.3381465.

[40] F. A. Shah, K. Sirts, D. Pfahl, F. A. S. B, K. Sirts, and D. Pfahl, "Is the SAFE Approach Too Simple for App Feature Extraction? A Replication Study," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1, Springer International Publishing, 2019, pp. 21–36. doi: 10.1007/978-3-030-15538-4_2.

[41] N. Z. Dina and N. Juniarta, "Deriving Customers Preferences for Hotels From Unstructured Data," *Geojournal of Tourism and Geosites* , vol. 43, no. 3, pp. 872–877, 2022, doi: 10.30892/gtg.43305-899.

[42] J. An, H. Cho, H. Kwak, M. Z. Hassen, and B. J. Jansen, "Towards Automatic Persona Generation Using Social Media," in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, IEEE, Aug. 2016, pp. 206–211. doi: 10.1109/W-FiCloud.2016.51.

[43] S. G. Jung, J. An, H. Kwak, M. Ahmad, L. Nielsen, and B. J. Jansen, "Persona generation from aggregated social media data," in *Conference on Human Factors in Computing Systems - Proceedings*, 2017, pp. 1748–1755. doi: 10.1145/3027063.3053120.

[44] S. gyo Jung, J. Salminen, H. Kwak, J. An, and B. J. Jansen, "Automatic Persona Generation (APG): A rationale and demonstration," in *CHIIR 2018 - Proceedings of the 2018 Conference on Human Information Interaction and Retrieval*, 2018, pp. 321–324. doi: 10.1145/3176349.3176893.

[45] G. Wu and J. Zhu, "Multi-label classification: do hamming loss and subset accuracy really conflict with each other?," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, in NIPS '20. Red Hook, NY, USA: Curran Associates Inc., 2020.