

**ORIGINAL RESEARCH**

# **USER STORY EXTRACTION FROM ONLINE NEWS WITH FEATURE-BASED AND MAXIMUM ENTROPY METHOD FOR SOFTWARE REQUIREMENTS ELICITATION**

Nafingatun Ngaliah<sup>1</sup> | Daniel Siahaan\*<sup>1</sup> | Indra Kharisma Raharjana<sup>1,2</sup>

<sup>1</sup>Dept. of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>2</sup>Dept. of Information System, Airlangga University, Surabaya, Indonesia

**Correspondence**

\*Daniel Siahaan, Dept of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Email: daniel@if.its.ac.id

**Present Address**

Gedung Teknik Informatika, Jl. Teknik Kimia, Kampus ITS Sukolilo, Surabaya 60111, Indonesia

**Abstract**

Software requirements query is the first stage in software requirements engineering. Elicitation is the process of identifying software requirements from various sources such as interviews with resource persons, questionnaires, document analysis, etc. The user story is easy to adapt according to changing system requirements. The user story is a semi-structured language because the compilation of user stories must follow the syntax as a standard for writing features in agile software development methods. In addition, user story also easily understood by end-users who do not have an information technology background because they contain descriptions of system requirements in natural language. In making user stories, there are three aspects, namely the who aspect (actor), what aspect (activity), and the why aspect (reason). This study proposes the extraction of user stories consisting of who and what aspects of online news sites using feature extraction and maximum entropy as a classification method. The systems analyst can use the actual information related to the lessons obtained in the online news to get the required software requirements. The expected result of the extraction method in this research is to produce user stories relevant to the software requirements to assist systems analysts in generating requirements. This proposed method shows that the average precision and recall are 98.21% and 95.16% for the who aspect; 87,14% and 87,50% for what aspects; 81.21% and 78.60% for user stories. Thus, this result suggests that the proposed method generates user stories relevant to functional software.

**KEYWORDS:**

Feature-Based Classification, Maximum Entropy, Online News, Requirements Elicitation, User Story

## 1 | INTRODUCTION

The media development is very rapid, especially the internet or online media, to deliver information to the public. One way of delivering information to the public is online news. The systems analyst can use the information related to the lessons learned in the news to obtain the necessary software requirements. An example sentence from online news, "The landing gear discovery comes a day after investigators found the flight data recorder, commonly known as a "black box," from Lion Air Flight JT610". From the example of a news sentence, we can obtain actual information used as a software requirement, namely "investigators find flight data recorder." The information can be extracted into user stories. These user stories can be used as software functionalities of the plane crash domain.

At the software development stage, the needs analysis is carried out. Collecting data from software requirements can be done in various ways, such as interviews with resource persons, questionnaires, analyzing problems in the current system, or comparing several similar applications. After the software requirements are collected and analyzed, they must be described and documented. The user story is one of the documentation of the software requirements used in the agile method<sup>[1]</sup>. In contrast to other models (e.g., UML diagrams), which are static, user stories are easy to adapt according to changing system requirements. In addition, user stories are also easily understood by end-users who do not have an information technology background. This is because the user story describes the system requirements in natural language. The user story is a semi-structured language, meaning that user stories are structured by following the syntax of standard feature writing in agile software development methods.

Based on a survey that has been conducted in a study, it is explained that most software developers make requirements engineering using informal language (e.g. natural language)<sup>[2]</sup>. This is considered easier to understand and communicate with stakeholders. With the growing popularity of agile methods, the creation of user stories is growing rapidly<sup>[3]</sup>. Requirements for a system can be extracted and reused to produce the same and newer systems. Research on feature extraction from natural language requirements has become one of the discussions in requirements engineering<sup>[4]</sup>.

This study proposes developing a user story extraction model with a feature extraction method and a maximum entropy classification model. The selection of features in the feature extraction process adjusts the online news dataset, namely the dependency parse, named-entity recognition (NER) or POS Tagging feature, and word-trigger. The maximum entropy trained model was used at the classification stage. This user story extraction model only focuses on two aspects of the user story, namely the aspect of who and the aspect of what. The selection of the two aspects is based on the format of the simplest user story (as a [who], I want [what]). The requirements were set as the candidates for the aspect of what and aspect of why. However, they have similar structures, i.e., verb + noun phrase. Therefore, our user story extraction would only focus on the aspect of who and aspect of what.

## 2 | PREVIOUS RESEARCHES

Previous research proposed a visual narrative approach that extracts conceptual models from user stories based on natural language processing. This approach was chosen because it is popular among practitioners who focus on agile methods and the essential components of the user story, namely who, what, and why<sup>[5]</sup>. Research in this area mainly collects data from software documentation, such as software reviews<sup>[6, 7]</sup> and requirement specification documents<sup>[8]</sup>. Meanwhile, research in the field of natural language that uses and processes web data to be converted into requirements is still rarely found.

Research on extraction methods to detect named entities, such as names of people, locations, organizations, etc., is a challenge for researchers<sup>[9]</sup>. There are three techniques for determining NER: a rule-based approach, a learning-based approach, and a hybrid approach. Learning-based systems include supervised, semi-supervised, and unsupervised learning<sup>[10]</sup>. NER research in the biomedical domain uses a supervised learning technique, namely maximum entropy classification. The maximum entropy or Maxent method has been widely used in natural language processing (NLP) classification because it has been proven to improve classification results<sup>[11, 12]</sup>. Maximum entropy is a classification method that classifies data based on the entropy value. Entropy is used to measure the heterogeneity or diversity of a data set. The entropy value also gets more significant if the data sample is more heterogeneous<sup>[13]</sup>. Research related to NER with maximum entropy classification proved to get good results<sup>[14]</sup>. Extraction of information in a social media application, namely Twitter, with experimental results showing the success of the

research method with a 90.3% precision value. Maximum entropy is the fastest method for extracting and classifying entity sets from the database compared to the Hidden-Markov model and SVM<sup>[12]</sup>.

In addition to using the maximum entropy classification, this research uses feature extraction and reduction techniques to obtain relevant results and sound system performance<sup>[12]</sup>. The use of words as triggers for feature extraction is widely used, especially in NER research<sup>[10, 14]</sup>. The two studies examined biomedical NERs, with one of the feature extractions used being word triggers. There are two types of word triggers: noun triggers and verb triggers<sup>[10]</sup>. Extracting word triggers consisting of head noun triggers and verb triggers based on the frequency of occurrence in the training dataset. Meanwhile, Shen et al.<sup>[15]</sup> extracts word-triggers based on the semantic approach of keywords. Keywords are obtained by extracting training data and taking 60% of the top word rankings that often appear. This study adopts the classification and feature extraction method<sup>[10]</sup> by adding the dependency parsing feature. The addition of this feature is by the need to extract user stories. In addition to the newly added features, the study does not use a reduction technique as in the study<sup>[10]</sup>. The feature extraction method and maximum entropy classification were selected to produce the relevance aspect of who as shown in previous studies related to NER.

The method proposed by Raharjana et al.<sup>[16]</sup> extracts information from online news to build into a user story. The requirement for online news information is to have elements of who, what, and why which can be compiled into a user story. The information retrieval was carried out using the NLP method and a case study regarding the earthquake in Palu. Retrieval of information from online news to be extracted into user stories has succeeded in obtaining 105 user stories. The resulting user stories include 41 aspects of who, 94 aspects of what, and eight aspects of why. In addition to the conceptual model for user story extraction, Raharjana et al.<sup>[17]</sup> also represents an SLR (Systematic Literature Review) implementation of NLP in the user story extraction process. One of the findings of SLR is that the POS tag is the most widely used NLP technique, so this study uses the POS tag as a feature in the extraction process of aspect of who and aspect of what. In addition to using POS tagging techniques, semantic approaches are also starting to gain a place in the research field<sup>[18]</sup>.

This study is part of the bigger research<sup>[16]</sup>. This study proposes a method for extracting user stories from online news sites using a syntactic approach. In this study, the authors propose extracting user stories consisting of aspects of who (actors) and what (activities) from online news sites. The selection of aspects is based on the preparation of the simplest user story, namely "As a [who], I want [what]," besides the requirements as candidates for the aspect of what and the aspect of why have similarities (consisting of a verb + noun phrase pattern), so that researcher chooses user story extraction which focuses on the aspects of who and what. The online news that will be used as user story extraction is online news with the plane crash domain. Text data on online news with an irregular structure is first identified as candidates for the who and what aspects. In general, online news sources are not related to software products, but online news contains information that can be analyzed and used as system requirements or features. Our method would help systems analysts elicit the software requirements of a software development project.

### 3 | MATERIAL AND METHOD

This chapter explains the process of developing the extraction method used in this study. The development of the extraction method is depicted in Figure 1 . The method development stage is divided into five parts: data preprocessing, feature extraction, maximum entropy modeling, maximum entropy classification, and evaluation or testing.

#### 3.1 | Online News Dataset

The data collected is from online news documents of a specific domain problem. The proposed method would process the online news and produce software requirements. The online news was obtained from trusted online news sites, such as CNN, BBC, Chanel News Asia, etc. The domain problem is related to handling aircraft accidents and their evacuation process. The news should be written in English and related to airplane accidents.

The stages in data collection are determining keywords in news collection, searching for news using the Google Search API with the query "airplane crash," downloading news search results with a crawling system, and sorting news. The data in the form of a collection of news is entered into a .txt file. Online news as the primary source influences the quality of the aspect of who and aspect of what is produced. News search queries are essential in determining the quality of the input dataset. Several news stories contain statements or sentences with political and economic elements from these news documents. Therefore, the last

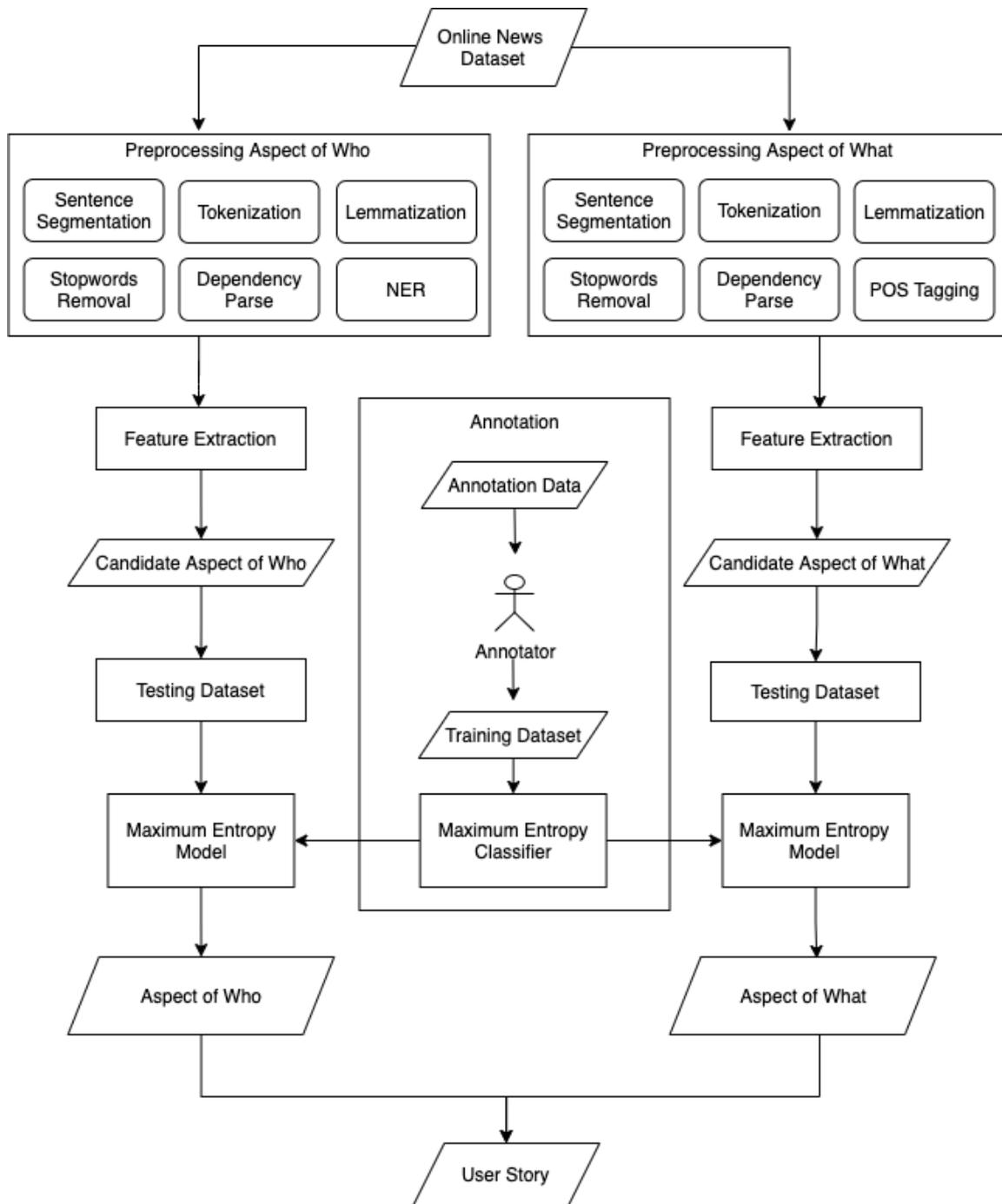


FIGURE 1 The sliced sample.

step is collecting and sorting the online news. In this study, the dataset of online news was crawled from three sources, i.e., BBC News, Channel News Asia, and CNN. There is 25 news (21 news for training and 4 news for testing ). There are 709 sentences in total. On average, there are 25 sentences in each news.

### 3.2 | Preprocessing

Each dataset to be feature extracted will go through preprocessing first. Preprocessing is carried out to get the data input format and eliminate words or punctuation that are not needed in the feature extraction process. The preprocessing includes tokenization, lemmatization, stopwords removal, dependency parse, NER, and POS Tagging. The input in this process is a sentence. For

example, the sentence "he says KNKT have yet to speak to Sriwijaya Air's management, but we collect data on the plane and the pilot." The resulting output consists of the dependency parsing, NER, and POS Tagging results. Given the previous sentence, dependency parsing would produce "he (nsubj), KNKT (nsubj), Sriwijaya Air's management (pobj), data (dobj), the plane (pobj), the pilot (conj)." The NER would produce "knkt (ORG), Sriwijaya Air (ORG)," and result of POS tagging is "he PRON, say VERB, KNKT PROPN, have VERB, yet ADV, to PART, speak VERB, to ADP, Sriwijaya PROPN, Air PROPN, management NOUN, but CCONJ, be VERB, collect VERB, data NOUN, on ADP, the DET, plane NOUN, and CCONJ, the DET, pilot NOUN."

### 3.3 | Annotation

Data annotation is the process of labeling data for use in the machine learning process. Data annotation in this study is the manual labeling stage on online news documents. This annotation aims to mark words or phrases from online news based on class features (aspect of who and aspect of what classes). Three annotators carried out data annotation. The input of this annotation process is the annotation data, while the output is the training dataset. The training dataset is used to train the maximum entropy classification model.

### 3.4 | Feature Extraction

After going through preprocessing, the phrases resulting from the preprocessing are extracted and used in the feature extraction process. The feature extraction results are the candidate aspect of who and the candidate aspect of what. The method could consist of dependency, post-tag, NER, and word-trigger modules. Dependency parse is used to find the types of noun chunks (nsubj, xsubj, and nsubjpass) used as candidates for the aspect of who. In addition, dependency parse also removes noun chunks with the lexnames "noun person." Meanwhile, in the search for candidate aspects of what, the dependency parse process is used to search for syntactic relationships between verb phrases (VERB) followed by noun phrases (pobj, dobj). The NER process aims to detect the organization's name (ORG) from the input sentences used as a candidate aspect of who.

POS tagging is used to get verb (VERB, VBN) and noun (NOUN) patterns for each word in a sentence. The POS tagging process is carried out because there are verbs that are not detected in the dependency parsing process. For example, in the sentence "the police hospital have taken 40 DNA samples from the relative of victims and other medical records to help with identification, official say", the dependency parsing would produce the following aspect of what : "take a DNA sample," while using POS tagging, it would produce "take DNA" and "help identification." From the example sentence, the result of verb + noun phrase in POS tagging is the same as the result of verb + noun phrase in dependency parse, so the result in the same POS tagging would be deleted. Unlike feature extraction in the dependency, post-tag, and NER features obtained from the previous preprocessing, word-trigger is obtained by searching for nouns and verbs that often appear in the annotation data. Examples of trigger nouns used are an agency, team, pilot, investigator, official, diver, and police. While examples of trigger verbs are search, identification, help, find, recover, take, and receive. Candidates from each aspect of the user story are then entered in a data frame format to simplify adding or modifying data. This data frame contains sentences, sentence indexes, and each candidate aspect of the user story.

### 3.5 | Maximum Entropy Classification

Before the maximum entropy classification process, the maximum entropy classifier model is built using the maximum entropy algorithm. The construction of this classifier model uses a training dataset derived from annotated data that has been processed. The classifier model is useful for predicting the class from testing datasets. The maximum entropy classification algorithm in this process uses a module from NLTK, namely 'nltk.classify.maxent' with parameter algorithm='gis', trace=0, max\_iter=100. Algorithm selection was based on better results with the GIS algorithm than with the IIS algorithm. The maximum entropy model is then measured on how successfully the model performs the classification correctly using a testing dataset. The testing dataset obtained from preprocessing is an online news document different from the online news document used in the training dataset.

The classification process using maximum entropy is based on information from an online news sentence feature. The principle of the maximum entropy method is to find a probability distribution that gives the maximum entropy value. The Maximum Entropy model's probability distribution is similar to the training dataset distribution. The data frame of each aspect of the user

**TABLE 1** Example of class labeling results on aspect of who.

Sent ID	Sentence	Candidate Aspect of Who	Proability Distribution		$a^*$	Class
			Who	Non-Who		
7	body bag fill with human remain be spray with disinfectant at Jakarta's main port before be take to a police hospital	body bag	0.02	0.98	1	non-who
8	the police hospital have taken 40 DNA samples from the relative of the victim and other medical records to help with identification, official say.	police hospital official	0.91 0.48	0.09 0.52	0 1	who non-who
11	he says KNKT have yet to speak to Sriwijaya Air's management, but we collect data on the plane and the pilot.	knkt sriwijaya air	0.99 0.67	0.01 0.33	0 0	who who

**TABLE 2** Example of class labeling results on aspect of what.

Sent ID	Sentence	Candidate Aspect of Who	Proability Distribution		$a^*$	Class
			Who	Non-Who		
7	body bag fill with human remain be spray with disinfectant at Jakarta's main port before be take to a police hospital	fill human remain spray disinfectant take police	0.12 0.33 0.24	0.88 0.67 0.76	1 1 1	non-what non-what non-what
8	the police hospital have taken 40 DNA samples from the relative of the victim and other medical records to help with identification, official say.	take dna sample help identification	0.60 0.85	0.40 0.15	0 0	what what
11	he says KNKT have yet to speak to Sriwijaya Air's management, but we collect data on the plane and the pilot.	speak management collect data	0.44 0.93	0.56 0.07	0 0	what what

story is entered into a dictionary, items, and keys following the format in the training dataset. Then the probability distribution is calculated using the function from nltk, namely 'classifier.prob\_classify'.

The probability distribution results then determine the class label for each aspect of the user story with the equation (1). Sentence with sent id 11 in Table 2 with candidate aspect of what "speak management," the determination of class labels is calculated using Eq. 1.

$$a^* = \arg \max_{a \in A} p(a, b) \quad (1)$$

$$a^* = \arg \max_{a \in A} p(0.43609, 0.56931)$$

$$a^* = 1$$

The calculation shows that  $a^* = "1"$ , which means that the class label on the candidate aspect of what "speak management" is non-what. If the result is  $a^* = "0"$ , then the class label is who. Thus, we see the examples of class labeling results in Table 1 (for the aspect of who) and Table 2 (for the aspect of what).

### 3.6 | Evaluation

This research will test three scenarios in the testing process: the aspect of who scenario, the spect of what scenario, and the user story scenario. Testing is carried out on testing datasets to measure the success of the maximum entropy model by looking for accuracy, precision, and recall values. The testing dataset used is four news stories with 180 sentences. Apart from these scenarios, tests were also carried out on different architectures to see the difference in the extraction results. The architectures include only dependency parsing, merging dependency parse and NER or POS Tagging, and merging dependency parse, NER, or POS Tagging, and word triggers).

To evaluate the results of the proposed method in this study, we analyzed the precision, recall, and f-measure values based on the extraction results and ground-truth of each test scenario. ground-truth from the aspect of who, aspect of what, and user story

**TABLE 3** Performance metrics on aspect of who.

Feature	Performance Metrics				#Aspect
	Accuracy	Precision	Recall	F-Measure	
Dependency	31.98%	65.10%	51.80%	26.67%	138
Dependency, NER	88.84%	94.21%	62.12%	66.44%	32
Dependency, NER, Word-Trigger	97.32%	98.24%	95.00%	96.47%	60

**TABLE 4** Example results of method answers and ground-truth answers on aspects of who.

Sentence	Result of Answer Method			Ground Truth
	1-feature	2-feature	3-feature	
Indonesian navy divers scouring the seabed on Monday (Jan 11) were closing in on data recorders from a Sriwijaya Air jet that crashed at the weekend, as investigators took up the grim task of identifying victims mangled remains.	indonesian navy divers investigators	-	indonesian navy divers investigators	indonesian navy divers -
Finding the boxes - cockpit voice and flight data recorders - could offer crucial clues as to why the Sriwijaya Air Boeing 737-500 plunged about 10,000ft in less than a minute before slamming into the Java Sea.	-	-	-	-
Divers had further narrowed down the search area strongly suspected to be the location of the black boxes, a Navy spokesman, Fajar Tri Rohadi, said	divers -	navy -	divers navy	divers -
Bayu Wardoyo, a rescue diver, told Kompas TV more human remains had been found on Monday and that underwater visibility was about 5m to 6m	bayu wardoyo	kompas tv	-	-

in each online news sentence were obtained from the recommendations by three experts. In addition, the performance of this proposed method is also measured by calculating the agreement value of the three annotators using the Kappa method.

### 3.7 | Testing Scenario

The steps used in this scenario are extracting the aspect of who and aspect of what from online news with the proposed method, then the results of the two aspects are compiled into a user story manually. Extraction is done using three extraction features, namely dependency parse, NER or POS Tagging, and word-trigger. Then, the measurement matrix values are calculated from the extraction results, including precision, recall, accuracy, and f-measure. The measurement matrix value of each extraction result is based on comparing one or more extraction features.

#### 3.7.1 | Scenario I (Aspect of Who)

The aspect of who extraction test obtained results as in Table 3 . The best results were using three features with 97.32% accuracy, 98.24% precision, 95.00% recall, and 96.47% f-measure and total aspect of who—obtained as many as 60. The value of the aspect of who matrix using (1) feature, (2) features, and (3) features has a significant difference. The more features used, the more heterogeneous the probability distribution value. If the probability distribution is more heterogeneous, the maximum entropy model will be more precise in class labeling. Significant differences in inaccuracy results from the one feature to two features also occur because the noun phrase results in the first feature would be eliminated in the second feature method if a noun phrase result has a label other than "ORG." Hence, the aspect of who produced is more dominant in the noun phrase, which has the label "ORG" and causes the number of aspects of who to decrease significantly.

Meanwhile, the difference in the accuracy of the second and third architectures is insignificant because adding the word-trigger feature would only increase the probability distribution value of the noun phrase. Furthermore, the aspect of who was not dominated by the noun phrase produced merely by NER and the other modules. features to 3 features also increase.

Moreover, Table 4 shows the example of extracting the aspect of who is using each architecture compared to the ground-truth. In sentence 1, the first and third architecture could produce a pretty similar aspect of who with ground-truth. In method two, the aspect of who feature is dominated by a noun phrase with the label NER "ORG." Then in sentence 2, the three combined features do not produce the aspect of who. In sentence 3, the aspect of who ('drivers') produced by the first architecture was initially a

**TABLE 5** Performance metrics on aspect of what.

Feature	Performance Metrics				#Aspect
	Accuracy	Precision	Recall	F-Measure	
Dependency	9.44%	54.29%	50.52%	8.92%	91
Dependency, NER	68.38%	84.07%	51.14%	42.75%	89
Dependency, NER, Word-Trigger	85.48%	87.14%	87.50%	85.48%	112

**TABLE 6** Example results of method answers and ground-truth answers on aspects of what.

Sentence	Result of Answer Method			Ground Truth
	1-feature	2-feature	3-feature	
Indonesian navy divers scouring the seabed on Monday (Jan 11) were closing in on data recorders from a Sriwijaya Air jet that crashed at the weekend, as investigators took up the grim task of identifying victims' mangled remains	scour seabed - take grim task identify victims mangle remain	scour seabed close data take grim task identify victims mangle remain	scour seabed close data take grim task identify victims mangle remain"	scour seabed close in on data recorders take grim task identify victims mangle remain
Finding the boxes - cockpit voice and flight data recorders - could offer crucial clues as to why the Sriwijaya Air Boeing 737-500 plunged about 10,000ft in less than a minute before slamming into the Java Sea	find box cockpit voice flight data recorders  offer crucial clue	find box cockpit voice flight data recorders  "offer crucial clue"	find box cockpit voice flight data recorders "offer crucial clue"	find box cockpit voice and flight data recorders  offer crucial clue
Divers had further narrowed down the search area strongly suspected to be the location of the black boxes, a Navy spokesman, Fajar Tri Rohadi, said	narrow search area  -	narrow search area  -	narrow search area  -	narrow down the search area  find human remains
Bayu Wardoyo, a rescue diver, told Kompas TV more human remains had been found on Monday and that underwater visibility was about 5m to 6m	- -	find underwater visibility	find underwater visibility	find underwater visibility.

noun phrase with the type of dependency "nsubj." While in the second architecture, it ('navy') was initially a noun phrase with the label NER "ORG." Finally, in the third architecture, they ('drivers' and 'navy') were initially noun phrases of the two types of labels. Furthermore, in sentence 4, the aspect of who ('bayu wardoyo') produced by the first architecture was initially a noun phrase with the type of dependency "nsubj." While in the second architecture, even though 'bayu wardoyo' was a noun phrase, it was removed because the label NER is "PERSON." However, 'Kompas tv' was included because the label NER is 'ORG.'

### 3.7.2 | Scenario II (Aspect of What)

Testing on the extraction aspect of what obtained results as in Table 5, the best results were using three features with 85.48% accuracy, 87.14% precision, 87.50% recall, 85.48% f-measure, and the number of aspects of what obtained as many as 112. The value of the aspect of what matrix using each architecture has a significant difference. This is also due to the more features used, the more heterogeneous the probability distribution value. If the probability distribution is more heterogeneous, the maximum entropy model will be more precise in class labeling. A significant difference in the results of the matrix values using the first architecture and the second architecture occurs. This is because the extraction with the second architecture results in twice as many aspects of what candidates as compared to the first architecture. Nevertheless, the aspect of what results are more in the one feature method compared to 2 features. While the difference in accuracy results for the two features to the three features occurs because the addition of the word-trigger feature will increase the probability distribution value of the aspect of what candidate that has a verb-trigger so that the number aspect of what results also increases.

Table 6 illustrates the extracting aspect of what using each architecture in comparison with the ground-truth. In sentences 1 and 4, the second and third architecture produced similar aspects of what. However, different results were produced by the first architecture. This is because the phrase "close data" comes from the formation of POS tagging divided into word tokens so that the phrase formation for the candidate aspect of what becomes less than perfect. The difference between the two results is also found in sentence 3, where the extraction of the aspect of what from the proposed method is "narrow search area," while the ground-truth answer is "narrow down the search area." While in sentence four aspects of what from ground-truth, namely

**TABLE 7** Performance metrics on aspect of user story.

Feature	Performance Metrics				#User #Story
	Accuracy	Precision	Recall	F-Measure	
Dependency	71.78%	72.49%	71.38%	71.93%	91
Dependency, NER	72.68%	73.22%	72.44%	72.83%	89
Dependency, NER, Word-Trigger	78.20%	81.21%	78.60%	79.88%	112

**TABLE 8** Example results of method answers and ground-truth answers on user story.

Sent ID	Method			Ground-Truth		
	What	Who	UserStory	What	Who	UserStory
0	-	-	-	divers	close in on black boxes	As divers, I want to close in on black boxes
1	indonesian navy divers	scour seabed	As an indonesian navy diver, I want to scour the seabed	indonesian navy divers	scour seabed	As an indonesian navy diver, I want to scour the seabed.
	-	close data	As Indonesian navy divers, I want close data	-	close in on data recorders	As Indonesian navy diver, I want to close in on data recorders
2	investigators	take grim task	As investigator, I want to take grim task	investigators	take grim task	As investigator, I want to take grim task
	-	identify victims mangle remain	As investigator, I want to identify victims mangle remain	-	identify victims mangle remain	As investigator, I want to identify victims mangle remain
	-	find box cockpit voice flight data recorders	As user, I want to find box cockpit voice flight data recorders	-	find box cockpit voice and flight data recorders	As user, I want to find box cockpit voice flight data recorders
3	-	offer crucial clue	As user, I want to offer crucial clue	-	offer crucial clue	As user, I want to offer crucial clue
4	divers	narrow search area	As divers, I want to narrow search area	divers	narrow down the search area	As divers, I want to narrow search area
5	navy	-	-	-	-	-
	-	find underwater visibility	As user, I want underwater visibility	-	find human remains find underwater visibility	As divers, I want to find human remains As user, I want underwater visibility

"find human remains" are not found in the proposed method, this is because the phrase "human remains" has the dependency "nsubjpass" which is a candidate aspect of who.

### 3.7.3 | Scenario III (User Story)

Tests on user stories compiled manually from the extraction results of aspect of who and aspect of what can be seen in Table 7. The best results are using three features with 78.20% accuracy, 81.21% precision, 78.60% recall, f-measure 79.88%, and the number of user stories obtained is 112. The value of the user story matrix has a significant difference from the two aspects that compose it. For example, in the first method, the matrix value feature of accuracy, precision, and recall from the aspect of who is 31.98%, 65.10%, and 51.80%. In comparison, the matrix values of the aspect of what is 9.44%, 54.29%, and 50.52%, but the results of the user story matrix values are 71.78%, 72.49%, and 71.38%. This is because, in the preparation of user stories, the results from both aspects are combined, resulting in more candidate user stories. An example of combining the two aspects to get a candidate user story can be seen in Table 8.

Combining the two aspects of the user story, the confusion matrix is calculated for one feature, two features, and three features. The confusion matrix produces the user story measurement matrix (Table 7). The accumulation of the method with the first architecture resulted in 237 candidate user stories with 91 user stories and 146 non-user stories in the method answers, and ground-truth answers resulted in 80 user stories and 157 non-user stories. The method with two features resulted in 209 candidate user stories with 89 user stories and 120 non-user stories. The method and ground-truth answers yielded 80 user stories and 129

non-user stories. The method with three features resulted in 225 candidate user stories with 112 user stories and 113 non-user stories. The method and ground-truth answers yielded 80 user stories and 145 non-user stories.

### 3.8 | Result Analysis

The performance of the maximum entropy model approach is based on the suitability of the features used in the feature extraction<sup>[10]</sup>. Testing by comparing the proposed features. From these tests, the performance of the maximum entropy model of the first architecture resulted in a precision value of 64.76%, recall 66.85%, and f-measure 65.79%. In comparison, the combination of the second architecture resulted in a precision value of 67.52%, recall of 66.90%, and f-measure of 67.21%. In contrast, the third architecture produced a precision value of 67.86%, recall of 66.94%, and f-measure of 67.41%. The third architecture produces the best performance of the three combined features models.

While the extraction of aspect of who and aspect of what the best results are also obtained from the third architecture. The aspect of who obtained a precision value of 98.21%, recall 95.16%, and f-measure 96.55%, while the aspect of what obtained a precision value of 87.14%, recall 87.50%, and f-measure 85, 48%. From these measurements, the results obtained are 60 aspects of who and 112 aspects of what when combining three features. From the extraction with the best results, namely the three features method, a user story relevant to the software needs is in Table 8. The example in the user story sentence 1, namely "As Indonesian navy divers, I want to scour seabed," can be used as a software requirement by needs analysis, namely information regarding aircraft search operations and aircraft accident victims carried out by Navy divers.

## 4 | CONCLUSION

An investigation of water absorption for concrete using supplementary materials has been carried out. Based on the test result, there are five conclusions that can be drawn. First, in this case, supplementary materials, fly ash, and silica fume greatly impact the compressive strength test and water absorption test. Second, the optimum compressive strength was achieved by concrete with 15% fly ash as cement replacement. Third, the optimum sorptivity value was achieved by concrete with 15% fly ash as cement replacement. Fourth, when both fly ash and silica fume are adopted as a cement replacement, the sorptivity and compressive strength are better than just using silica fume. Finally, the water penetration becomes saturated on day 6th for every sample.

## CREDIT

**Nafingatun Ngaliah:** Methodology, Writing – original draft preparation, Formal analysis, Resource, Acquisition, and Investigation. **Daniel Siahaan:** Conceptualization, Supervision, and Writing - reiew and editing. **Indra Kharisma Raharjana:** Writing - review and editing, Data Acquisition, Resources.

## References

1. Schon EM, Thomaschewski J, Escalona MJ. Agile requirements engineering: a systematic literature review. *Computer Standards and Interfaces* 2017 1;49:79–91.
2. Neill CJ, Laplante PA. Requirements engineering: the State of the practice. *IEEE Software* 2003;20:40–45.
3. Kassab M. The changing landscape of requirements engineering practices over the past decade. In: 5th International Workshop on Empirical Requirements Engineering, EmpiRE 2015 Institute of Electrical and Electronics Engineers Inc.; 2016. p. 1–8.
4. Bakar NH, Kasirun ZM, Salleh N. Feature extraction approaches from natural language requirements for reuse in software product lines: A systematic literature review. *Journal of Systems and Software* 2015 8;106:132–149.
5. Lucassen G, Robeer M, Dalpiaz F, van der Werf JMEM, Brinkkemper S. Extracting conceptual models from user stories with visual narrator. *Requirements Engineering* 2017 9;22:339–358.

6. Bakar NH, Kasirun ZM, Salleh N, Jalab HA. Extracting features from online software reviews to aid requirements reuse. *Applied Soft Computing Journal* 2016 12;49:1297–1315.
7. Yang L, Liu B, Lin H, Lin Y. Combining local and global information for product feature extraction in opinion documents. *Information Processing Letters* 2016 10;116:623–627.
8. Sibarani EM, Hutagaol A, Simarmata D, Manihuruk J. Actor and Use Case Extraction from Text-Based Requirement Specification. In: *Proceedings of the International Conference Image Processing, Computers and Industrial Engineering (ICICIE'2014)*, Kuala Lumpur, Malaysia *Computers and Industrial Engineering*; 2014. p. 15–16.
9. Goyal A, Gupta V, Kumar M. Recent named entity recognition and classification techniques: A systematic review. *Computer Science Review* 2018 8;29:21–43.
10. Saha SK, Sarkar S, Mitra P. Feature selection techniques for maximum entropy based biomedical named entity recognition. *Journal of Biomedical Informatics* 2009 10;42:905–911.
11. Putra TSJ. *A Maximum Entropy Approach to Named Entity Recognition*. PhD thesis, New York University; 1999.
12. Swain C, Bansod PP, Janghel S. Name entity recognition by using maximum entropy. *International Journal of Applied Engineering Research* 2013;8:2241–2244.
13. Syah AP, Adiwijaya, Faraby SA. Analisis Sentimen Pada Data Ulasan Produk Toko Online Dengan Metode Maximum Entropy Sentiment Analysis on Online Store Product Reviews With Maximum. In: *Proceeding of Engineering ( E-Proceeding) Proceeding of Engineering ( E-Proceeding)*; 2017. p. 4632–4640.
14. Jung JJ. Online named entity recognition method for microtexts in social networking services: A case study of twitter. *Expert Systems with Applications* 2012 7;39:8066–8070.
15. Shen D, Zhang J, Zhou G, Su J, Tan CL. Effective Adaptation of a Hidden Markov Model-based Named Entity Recognizer for Biomedical Domain. In: *Proceedings of the ACL 2003 workshop on Natural language Processing in Biomedicine Proceedings of the ACL 2003 workshop on Natural language Processing in Biomedicine*; 2003. p. 49–56.
16. Raharjana IK, Siahaan D, Faticah C. User Story Extraction from Online News for Software Requirements Elicitation: A Conceptual Model. In: *JCSSE 2019 - 16th International Joint Conference on Computer Science and Software Engineering: Knowledge Evolution Towards Singularity of Man-Machine Intelligence JCSSE 2019*; 2019. p. 342–347. <https://scholar.unair.ac.id/en/publications/user-story-extraction-from-online-news-for-software-requirements->
17. Raharjana IK, Siahaan D, Faticah C. User stories and natural language processing: a systematic literature review. *IEEE Access* 2021;9:53811–53826.
18. Dewi MR, Raharjana IK, Siahaan D, Faticah C. 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: *10th International Conference on Software and Computer Applications 10th International Conference on Software and Computer Applications*; 2021. p. 81–87.

**How to cite this article:** Ngaliah N., Siahaan D., Raharjana I.K. (2021), User Story Extraction from Online News with Feature-Based and Maximum Entropy Method for Software Requirements Elicitation, *IPTEK The Journal of Technology and Science*, 32(3):125-135.