

Korean News Text Summarizer Enriched with Major Information Items

Meiying Ren¹ and Sinjae Kang*

Department of Computer & Information Engineering, Daegu University
¹meeyeong1211@gmail.com, *sjkang@daegu.ac.kr

Abstract

In news text summarization, one commonly used method is extractive summarization. However, extractive summarization has the probability that some major information, like time or location, could be missing. This paper presents a method that tries to extract who, where, when, and a proposed predicate element (PE) as major information items (MIIs) that are considered core information but are easily lost in extractive news summarization. First, we extracted MIIs, and then enriched extracted summaries by adding these elements to them. Previous studies used a rules-based method to extract the 5W1H elements (who, what, where, when, why, and how) and utilized them in weighting sentences in order to extract a summary. However, this still does not address the problem if a selected sentence does not contain core information. The method in this study first conducted filtering using some high-precision rules. Then, conditional random field-based labeling was performed to extract MIIs from the news text. Afterwards, we conducted extractive summarization and added the MIIs.

Keywords: major information item, extractive summarization, hybrid approach, machine learning

1. Introduction

Nowadays, because of the tremendous increase in Internet documents, the requirement for concise information is growing. Therefore, text summarization is not only one of the research areas, but it has also become a necessary commercial service. Internet news is one field that needs the service, especially since most news content these days is consumed online, so quality of service (QoS) would increase if a correct and concise summary is provided alongside. There are two techniques in automatic text summarization: abstractive summarization and extractive summarization.

Abstractive summarization understands the content of the document first, and then generates new summary sentences, just as human beings do. The generated summary is natural, although generating sentences without grammatical and semantic errors is not a simple matter.

Extractive summarization has been explored much more than abstractive summarization. Extractive summarization selects sentences containing important information and combines them to obtain a summary text. Using this method, the summary sentences themselves would be natural, but could lack coherence because the sentences come from different parts of the news article. Additionally, because the approach summarizes the document according to a fixed length or ratio, information that is considered important could be missed [1].

Therefore, this study tries to enrich the information of a summary by using major information items extracted beforehand. Commonly, the major information items

Received (October 21, 2017), Review Result (January 25, 2018), Accepted (February 2, 2018)

* Corresponding Author

considered important in news articles are the 5W1H items: who, what, where, when, why, and how. But the *what* of an item is usually contained in the title and first paragraph, and we consider *why* and *how* as additional information. So we selected *who*, *where*, and *when* as major information items, since they are easily missed. Furthermore, we add a new *predicate element* (PE), which is the main verb that indicates the major action of the summary. We extract these information items and include them in the extracted summary text so they can enrich the summary text.

This paper explains extracting major information items (MIIs) using a hybrid method combining rules-based and conditional random field (CRF) methods. Afterwards, we combine the MII extraction system with the extractive summarization system to obtain better summaries. We used 1,000 manually tagged news documents and used Bilingual Evaluation Understudy (BLEU) and Recall-Oriented Understudy for Gisting Evaluation (L) (Rouge-L) scores as the evaluation method.

2. Literature Review

2.1. Online News Documents

Important information in news items is usually disseminated in the title and the first paragraph, then in the body text in order of importance, which is called the *inverted pyramid* style. According to Brandow *et al.*, [2] and Okumura *et al.*, [3], 60% of information in a news article is contained in the title and the first paragraph, and the lead-based summarization methodology shows performance accuracy of around 60% as well. Considering that news content is often consumed online, online news may contain meaningless words or overly exaggerated or implied expressions simply to increase the number of views. In these situations, some redundant information could be included in both the title and the first paragraph. Figure 1 shows an example. The news item is mainly about a Korean sports star who “Liked” a post by a Russian sports star who is a rival of the nation’s most-loved skater. However, the title is “*Shocking, Sohn should naturalize to Russia? ...*,” which does not reflect the actual news.



Figure 1. Sample News Data

(from: <http://www.simnilbo.co.kr/news/articleView.html?idxno=549858>)

2.2. Text Summarization

2.2.1. Abstractive Summarization: Abstractive summarization creates a smooth and natural summary based on information extracted from documents. However, generating sentences, especially Korean sentences, is complicated. There have been some approaches

conducted by linguists [4]. However, for now, a lot of research in the language engineering field uses machine learning techniques, especially deep learning. The technique often used in abstractive summarization is the recurrent neural network (RNN) encoder-decoder model [5]. There were many studies done for the English language [5, 6, 7], but not many for Korean [8]. However, a well-known shortcoming of the RNN model is that it does not work properly when the input length is large [9, 10]. Data of the previously tried method for the Korean language was all short text [8]. And the methodology requires great amount of data to train the models.

2.2.2. Extractive Summarization: Extractive summarization creates a summary by selecting sentences from inside the document based on sentence scores calculated by capturing various features. Sentences in an extractive summary are natural, because they are written by humans, but the entire summary text could be abnormal because of incoherence between the sentences. The system extracts sentences with minimum duplication and maximum information. Some major information that is contained in other sentences could be missing. Features that are usually used in sentence selection are the sentence position, noun numbers in the sentence, and so on. Sentence selection algorithms commonly used include TextRank [11] or clustering.

2.3. Information Extraction from News Text

In news articles, the information commonly considered the most important is the 5W1H elements. These are considered the core information in formal text, such as news articles. In the Chinese domain, one study used semantic role labeling to extract the 5W1H elements [12]. They used a large amount of semantically labeled data to conduct the experiments. Sharma *et al.*, [13] described an extended approach to applying the method to English.

Within the Korean domain, Yoon *et al.*, [14] extracted and utilized the 5W1H elements to select sentences in an extractive text summarization system. Their method uses various knowledge bases, such as a thesaurus.

Considering that MIIs are information complementary to the summary, we selected information on *who*, *where*, and *when*, because these items can be distributed in sentences other than the selected ones. Moreover, referring to previous studies [12, 14], we added the *predicate element*, which is the verb used in the predicate of the summary text. Our system extracted *who*, *where*, *when*, and the *PE* as the major information items, and added them directly to the extracted summary to enrich the summary text.

3. Korean News Summarization System

3.1. Methodology

To construct our system, we first determined the major information items in the news article that could be missed. Among the 5W1H elements, *what* appears in almost every lede, and *why* and *how* can be seen as additional information. Therefore, we selected *who*, *where*, and *when* as major information. Additionally, we include the PE, since we found the main verb can sometimes be hard to decide upon, due to the complexity of the Korean verb system. Our goal is to use these items to enrich the extracted summaries.

We performed preprocessing steps first: named entity recognition (NER) and co-reference resolution. Figure 2 shows the preprocessing steps. Subsequently, major information items are recognized by following the corresponding rules for each element. At the same time, with the preprocessed data, we conduct sentence scoring and sentence clustering.

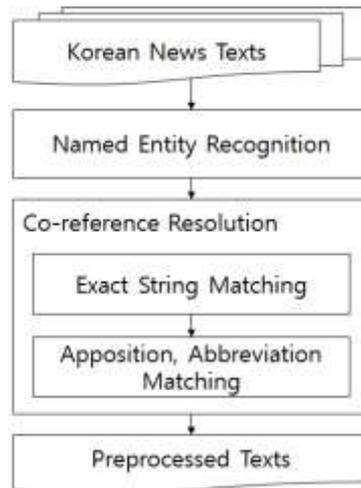


Figure 2. Preprocessing Korean News Texts

3.1.1. Proposed System: Figure 3 shows the overall process of the proposed system after preprocessing.

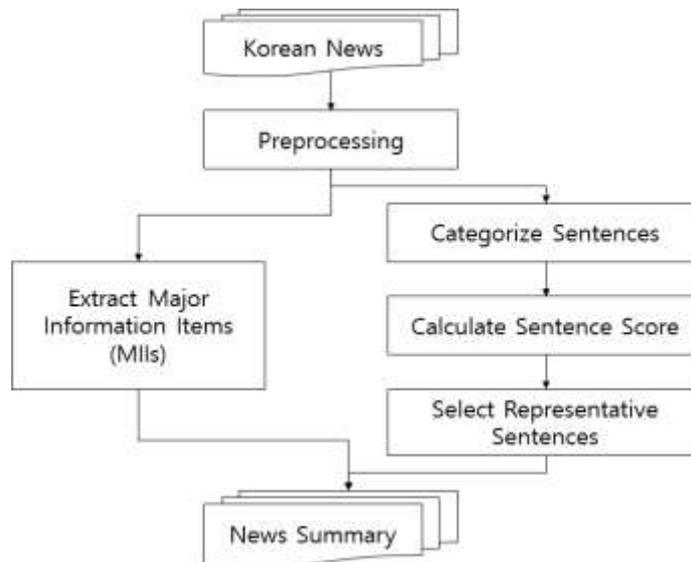


Figure 3. Overall Process of the Korean News Summarization System

To combine extracted MIIs, we need to extract a summary first. The system receives the preprocessed news as input. Then, we conduct sentence clustering using cosine similarity. The threshold was set at 0.3. Throughout this step, the sentences are categorized into several classes. Afterwards, we score sentences in each category based on some basic features, such as sentence position, the number of named entities, and the number of title words.

In the equation expressed below, we assume there is a certain number of sentences in document D , and the sentence index is noted as i . $Score_i$ is the total sentence score of sentence i . $Score_i^p$ is the position score of sentence i . $Score_i^{NER}$ is the NER score, while $Score_i^{TitleWord}$ is the title word score.

$$Score_i^p = \frac{Length_D}{i + 1}, i \geq 0 \quad (1)$$

$$Score_i^{NER} = \text{number of NER words} \quad (2)$$

$$Score_i^{TitleWord} = \text{number of title words} \quad (3)$$

$$Score_i = \sum(\log(Score_i^P), \log(Score_i^{NER}), \log(Score_i^{TitleWord})) \quad (4)$$

3.1.2. Extraction of Korean MIIs: The categories we referred to, which are from the 5W1H elements and are defined by the National Institute of Korean Language, are as follows [15].

- 1) Who: when a person, animal, country, or organization is the subject
- 2) When: time
- 3) Where: place name, agency
- 4) What: the object of the predicate
- 5) How: a predicate with a connective suffix ‘-아/-어/-다/-며 (-a/-e/-da/-myeo)’
- 6) Why: ‘-to do ...’, ‘-to ...’

We select *who*, *where*, and *when*, and add the *PE*, as explained before. The system recognizes key information with high accuracy through a rules-based system, and then, recognizes the remaining information through sequential labeling.

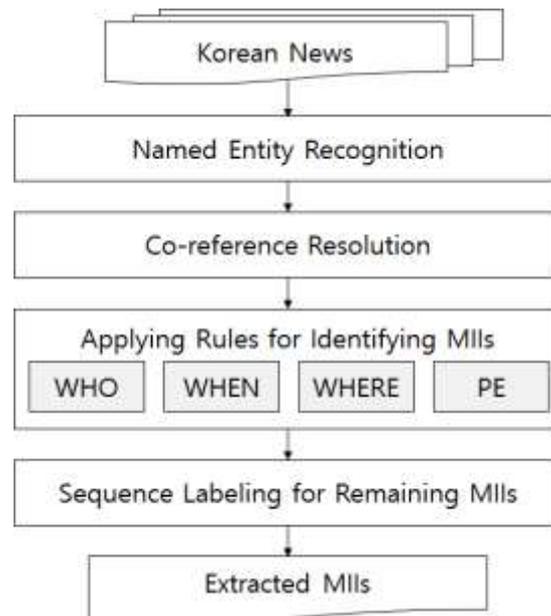


Figure 4. Process of the Korean MII Extraction Module

◆ The rules for [WHO] are as follows.

Rule 1. If the title structure is <A, ~>, assign A to [WHO] when A is a person / organization. For instance, in “이명박 대통령, 야당 대표와 11 일 이후 회동 추진 (President Lee Myung-bak will meet with opposition leader after 11th)”, “President Lee Myung-bak” should be allocated to WHO.

Rule 2. If ‘person/organization/country + 은/는(eun/nun)’ appears in the lede, assign person/organization/country to [WHO]. For instance, in “국가보훈처는 일제에 빼앗긴 국권을 되찾기 위해... (To reclaim the national sovereignty from Japanese imperialism, the Ministry of Patriots and Veterans Affairs is trying to ...)”, “the Ministry of Patriots and Veterans Affairs” should be allocated to [WHO].

Rule 3. If ‘person/organization/country + 가/이(ga/e)’ appears in the lede, allocate person/organization/country to [WHO]. For instance, in “이명박 대통령이 민주당 정세균 대표와의 회동을 추진하는 것으로 알려졌다.... (*President Lee Myung-bak is reportedly pushing for a meeting with Democratic Party leader Chung Sye-kyun....*)”, “*President Lee Myung-bak*” would be allocated to [WHO].

Rule 4. If ‘person/organization/country + 와/과(wa/gwa) + person/organization/country + 가/이(ga/e)’ appears in the lede, [WHO] will be both the first and the second named entities. For instance, in “한나라당 박희태 대표와 민주당 정세균 대표가 31 일 오후 양자회동을 갖고 국회 파국을 막아야 한다는 데 의견을 같이했다.... (*The Grand National Party chairman Park Hee-tae and Democratic Party leader Jeong Syeon-kyun met on the afternoon of January 31st, and agreed to stop the collapse of the National Assembly....*)”, “*The Grand National Party chairman Park Hee-tae and Democratic Party leader Jeong Syeon-kyun*” should be allocated to [WHO].

Rule 5. When Chinese characters are included in the title, replace them with the object indicated and assign it to [WHO]. For example, “與野” will be replaced with “ruling party and opposition party”.

◆ The rule for [WHEN] is as follows.

Rule 1. When there is only one time expression in the text, assign it to [WHEN]. For example, in the following named entity recognized news, there is only one time expression, which is “5 days”. Therefore, “5 days” is assigned to [WHEN].

한나라당/organization, 디도스 국민검증위원회/organization 인선회정 한나라당/organization 비상대책위원회/organization 는 5 일/time 국회 에서 전체회의를 열고 ... <i>The Grand National Party/organization, confirmed DiDos National People's Commitment Committee/organization The Grand National Party/organization's emergency committee/organization held a full meeting on the 5th /time...</i>

◆ The rules for [WHERE] are as follows.

Rule 1. Assign [location] to [WHERE] when the lede is a <~ [location]> ~ structure. For example, in the following news, assign "North Korea" to [WHERE].

北서 한마디씩 공개되는 ‘김정일 유언’ 북한/location 에서 김정일/person 국방위원장/title 의 ‘유언’이 잇따라 공개되고 있다. <i>Kim Jong Il's testament is released word for word in North Korea In North Korea/location, Kim Jong Il/person Chairman of the National Defense Commission/title's 'testaments' are being released ...</i>
--

Rule 2. Assign [location] to [WHERE] when the lede contains a [location] + 에 서/에(e so/e) pattern. For example, in the following news, assign "Kimgu Memorial Hall" to [WHERE].

中 상하이·충칭서도 임정 수립 94 주년 기념식 국가보훈처/organization 는 일제에 빼앗긴 국권을 되찾기 위해 1919 년/time 수립된 대한민국임시정부/organization 의 수립 제 94 주년 기념식을 연다고 밝혔다. 12 일/time 보훈처에 따르면 13 일/time 오전 10 시/time 서울 용산구 효창동 백범김구기념관/location 에서 정홍원/person 국무총리/title 를 비롯해 박승춘/person 국가보훈처장/title, 원로 독립 유공자와 유족 등 1000 여명이 참석하는 가운데 기념식이 거행된다. ... <i>Ceremony for the 94th anniversary of establishing the Shanghai and Chongqing Sodo Provinces in China The Ministry of Patriots and Veterans Affairs announced the opening ceremony of the 94th anniversary of</i>
--

the establishment of the Provisional Government of Korea, which was established in 1919 in order to regain the national rights lost to the Japanese occupation. According to the Ministry of Patriots and Veterans, the ceremony will be held at the Kingu Memorial Hall in Hyangchang-dong, Yongsan-gu, Seoul, where more than 1,000 people are attending, including Prime Minister Choe Hong-won, Park Seung-chun, ...

◆ The rule for the [PREDICATE ENTITY] is as follows.

Rule 1. Assign the predicate of the sentence that contains the most nouns in the title to PE. For example, in the following news, the verb in the second sentence is assigned to the PE because the second sentence contains more information in the title.

28 일/time 영결식... 김정은/person 곁에 누가 설까
김정은/person 의 사람들은 누구일까. 20 대 후반의 지도자 김정은/person 을 옹위해 국가를 꾸려 갈 권력층의 윤곽은 28 일/time 열리는 김정일/person 국방위원장/title 의 영결식에서 드러날 전망이다.
27 일/time ...

*Funeral in 28th/time ... who will be next to Kim Jung Eun/person
Who are the people of Kim Jung Eun/person? The outline of the power struggle to defend the leader Kim Jung Eun in the late 20s is expected to be revealed at the funeral of Kim Jong Il/person Chairman of the National Defense Commission/title. 27th/time ...*

The input newspaper articles go through the rules-based system first by applying all these rules. Since the elements that correspond to MIIs are not a single word but a phrase, the sequence labeling model CRF is applied. We used words and NER tags as features for CRF, and constructed each CRF classifier for each element.

As mentioned earlier, the importance of information in newspaper articles is highly influenced by the position of the words and phrases in the newspaper article. Therefore, when constructing the data for CRF learning, we also put a START tag that indicates the starting point of new information. The NER tag of this tag is set to O. In Table 1, the first column is the word, the second column is the NER tag, the third column is the correct answer, and the last column is the system predicted answer. The tag B means the start words of the corresponding element; the I tag is annotated, except for the start word, and the O tag reveals information that does not correspond to the element.

Table 1. Sample of CRF Data

Word	NER Tag	System Answer	Real Answer
면제해 (<i>exemption</i>)	O	O	O
줘야 (<i>be</i>)	O	O	O
이명박 (<i>Lee, Myung-bak</i>)	Person	B	B
대통령 (<i>President</i>)	Title	I	I
은 (<i>postposition</i>)	O	O	O
28 일 (<i>28th</i>)	Time	O	O
국민권익위원회 (<i>Anti-Corruption and Civil Rights Commission</i>)	Organization	O	O

The rules for adding extracted MIIs are simple. *When* can be added to the beginning of the summary with a comma. For example, if the extracted *when* is ‘30 일 (30th)’, then we can add to the summary ‘30 일, ...’. *Where* can be added to the summary with the postpositional particle ‘~에서 (e seo)’. For *who*, we can save the postpositional particle of *who* when extracting the MIIs. Then, we add it to the beginning of the summary after *when* and *where* elements, since *who* usually plays the role of subject, and subjects usually appear at the start of the sentence.

4. Experiments

4.1. Data and Implementation Environments

In order to extract MIIs and utilize them in the summary, we performed the tagging work manually with 1,000 Internet news items. For each item, we tagged the MIIs. The system was implemented using Python, and CRF ++ [16] was used as the CRF library. Of the news items, 900 were used for CRF learning, and the remaining 100 were the test data.

The summarization data were also annotated when tagging Korean MIIs. To evaluate the results, we implemented BLEU and Rouge-L tests.

4.2. Experimental Results

4.2.1. Korean MII Extraction

Table 2. Results of Korean MII Extraction

Who			Where		
Precision	Recall	F-Measure	Precision	Recall	F-Measure
0.83	0.75	0.788	0.83	0.75	0.788
PE			When		
Precision	Recall	F-Measure	Precision	Recall	F-Measure
0.96	0.57	0.71	0.83	0.262	0.398

As we can see from Table 2, *who* and *where* were extracted properly. *PE* was extracted with good precision but low recall, and *when* was extracted with even lower recall. The reason could be due to an annotation problem because of the complexity of the Korean verb system. For example, in “...보이지 않는다고 밝혀졌다(... *revealed cannot see*),” there are three verbs and sometimes *cannot see* seems important but sometimes *reveal* can be more important. According to the contextual differences, annotation can vary in different cases. For *when*, there are many cases where the news itself does not indicate when the news occurred.

4.2.2. Korean Extractive Summarization Enriched by Korean MIIs

Table 3. Results of the Korean Extractive Summarization System

Measurement Experiment	BLEU Score		Rouge-L Score
	Unigram	Bigram	
System Without MIIs	0.114	0.0973	0.1886
System With MIIs	0.1975	0.1773	0.3250

Table 3 shows that the system with MIIs has higher performance than the system without MIIs. Below are sample results.

Table 4. Sample Summarization Results

4-1) Sample news article

<p>안철수 “국민 눈높이 안맞는 인사기준부터 바꿔야”</p> <p>▲ 모두 발언하는 안철수 공동대표, 안철수 새정치민주연합 공동대표가 30 일 오전 국회에서 열린 최고위원회의에서 모두 발언을 하고 있다.</p> <p>새정치민주연합 안철수 공동대표는 잇단 총리 후보자 낙마 사태 이후 여권에서 인사청문회 제도개선을 추진하는 것과 관련, “제도 문제 이전에 사람의 문제”라면서 “국민 눈높이에 맞지 않는 대통령의 인사기준부터 바뀌어야 한다”고 말했다. 안 대표는 이날 국회에서 열린 최고위원회의에서 “문제는 국회에서 이뤄지는 청문회가 아니라 청문회 오기전 언론과 국민 검증에서 자격미달로 밝혀지는 것”이라며 이같이 밝혔다. 안 대표는 또 “대통령과 여당이 다수임을 믿고 청문회를 후퇴시키려 하거나 기준을 낮추려고 하면 여론 청문회는 더 가혹해질 것”이라면서 “자격이 안 되는 사람을 국회에서 다수로 밀어붙인다 해서 국민이 동의하지는 않을 것”이라고 지적했다. 김한길 공동대표는 “대통령은 적폐해소를 말하지만 인사청문회 대상의 여러 명이 적폐 그 자체”라면서 “문제 인사들을 내놓고 인사청문회 제도를 손보자는 주장은 입학시험에 자꾸 떨어지니 입시제도 바꾸자는 주장과 같다”고 비판했다.</p> <p>김 대표는 전날 여야가 한민구 국방장관 후보자 인사청문보고서를 채택한 사실을 언급하며 “우리 당은 무조건 발목잡는 청문회는 하지 않을 것이나 문제 인사들 문제는 국민 시각에서 야무지게 따지고 확인할 것”이라고 강조했다.</p>
<p><i>Ahn, Chul-soo, "We have to change the personal management standards that are not in the public eye."</i></p> <p>▲ <i>Ahn, Cheol-soo, the chairman of the New Political and Democratic Union, is making a speech at the top committee held on the morning of the 30th.</i></p> <p><i>Ahn, Cheol-soo, the co-chairman of the New Political Democratic Union, said, "Before a system problem, it is a human problem." "It should change from the president's HR management standards that are not in line with the public's expectations," he said, referring to the ruling party's promotion of the Confirmation Hearing System after the successive unseating of candidates for prime minister.</i></p> <p><i>On the day of the highest committee at the National Assembly, Ahn said, "The problem is not the Hearing in the National Assembly, but their qualification in the media and public verification before the Hearing." Ahn also said, "If the president and the ruling party believe that they are the majority and try to retreat or lower the Hearing, public hearings will become more severe." "The public will not agree to the person who is not qualified just because the majority in the National Assembly is pushing", he pointed out. Kim Han-gil, chairman of the joint committee, said, "The president says root out the accumulated corruption, but actually the target Confirmation Hearing is the deep-rooted corruption itself." Kim said, "Claiming to change the Confirmation Hearing system is like claiming to change the admission system because of continuous drops in the entrance exam." Kim spoke of the fact that Defense Minister Kim Han-gu candidate's Confirmation Hearing Report was adopted the previous day, and emphasized that, "Our party will not make a caviling Hearing, but the people who caused problems would be checked and judged in the view of public."</i></p>

4-2) Generated Summary Sample

Human summary	새정치민주연합 안철수 공동대표는 30 일 잇단 총리 후보자 낙마 사태 이후 여권에서 인사청문회 제도개선을 추진하는 것과 관련하여 “제도 문제 이전에 사람의 문제”라면서 “국민 눈높이에 맞지 않는 대통령의 인사기준부터 바뀌어야 한다”고 말했다.
Summary without MIIs	새정치민주연합 안철수 공동대표는 잇단 총리 후보자 낙마 사태 이후 여권에서 인사청문회 제도개선을 추진하는 것과 관련, “제도 문제 이전에 사람의 문제”라면서 “국민 눈높이에 맞지 않는 대통령의 인사기준부터 바뀌어야 한다”고 말했다. 안 대표는 이날 국회 에서 열린 최고위원회의에서 “문제는 국회 에서 이뤄지는 청문회 가 아니라 청문회 오기전 언론과 국민 검증에서 자격미달로 밝혀지는 것”이라며 이같이 밝혔다.
Summary of the proposed system	30 일 , 새정치민주연합 안철수 공동대표는 잇단 총리 후보자 낙마 사태 이후 여권에서 인사청문회 제도개선을 추진하는 것과 관련, “제도 문제 이전에 사람의 문제”라면서 “국민 눈높이에 맞지 않는 대통령의 인사기준부터 바뀌어야 한다”고 말했다. 안 대표는 이날 국회 에서 열린 최고위원회의에서 “문제는 국회 에서 이뤄지는 청문회 가 아니라 청문회 오기전 언론과 국민 검증에서 자격미달로 밝혀지는 것”이라며 이같이 밝혔다.
Human summary	<i>On the 30th, Ahn, Cheol-soo, the co-chairman of the New Political Democratic Union, said, "Before a system problem, it is a human problem." "It should change from the president's HR management standards that are not in line with the public's expectations," he said referring to the ruling party's promotion of the Confirmation Hearing System after the successive unseating of candidates for prime minister.</i>
Summary without MIIs	<i>The chairman of the New Political Democratic Union, Ahn Cheol-soo, said, "Before a system problem, it is a human problem." "It should change from the president's HR management standards that are not in line with the public's expectations," he said, referring to the ruling party's promotion of the Confirmation Hearing System after the successive unseating of candidates for prime minister. On the day of the highest committee at the National Assembly, Ahn said, "The problem is not the Hearing in the National Assembly, but their qualification in the media and public verification before the Hearing."</i>
Summary of the proposed system	<i>Ahn, Cheol-soo, a co-chairman of the new Political Democratic Union, said on the 30th, "After the fall of the Kim Dae-jung administration, we are trying to improve the human rights hearing system in the passport. Ahn said, "The problem is not a hearing in the National Assembly but a disqualification from the media and public verification before the hearing." On the day of the highest committee at the National Assembly, Ahn said, "The problem is not the Hearing in the National Assembly, but their qualification in the media and public verification before the Hearing."</i>

In Table 4-1), the sentence containing the time expression is a description of the picture. Therefore, the sentence would not be selected for the summary if we do not extract it.

Table 4-2) shows the generated sample summaries with and without MIIs. We can see in the proposed system, 30 ^일(30th) is added to the summary.

5. Conclusion and Future Research

In this study, we extracted key information from newspaper articles and applied it to extractive document summarization. The major information items we extracted were *who*, *where*, *when*, and the *PE*. We used a hybrid method to extract the MIIs, and added them to the extracted summary. The results showed some improvement, and this indicates that the method enriches summaries to some extent.

In the future, we intend to improve MII extraction performance and apply the key information to obtain abstractive summaries using deep learning.

Acknowledgments

This paper is a revised and expanded version of a paper entitled “Korean 5W1H Extraction Using Rule-based and Machine Learning Methods” presented at the 14th 2017 International Interdisciplinary Workshop, Daejeon, Korea, December 2017.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B03930346).

References

- [1] N. Bhatia and A. Jaiswal, “Trends in Extractive and Abstractive Techniques in Text Summarization”, International Journal of computer Applications., vol. 117, no. 6, (2015), pp. 21-24.
- [2] R. Brandow, K. Mitze and L. Rou, “Automatical Condensation of Electronic Publications by Sentence Selection”, Information Processing and Management., vol. 31, no. 6, (1995), pp. 675-685.
- [3] A. Okumura, T. Ikeda and K. Muraki, “Text Summarization Based on Information Extraction and Categorization Using 5W1H”, Journal of Natural Language Processing., vol. 6, no. 6, (1999), pp. 27-44.
- [4] G. Moon, “The Predication Operation and Multiple Subject Constructions in Korean”, Studies in Generative Grammar., vol. 10, no. 1, (2000), pp. 239-263.
- [5] R. Nallapati, B. Xiang and B. Zhou, “Sequence-to-sequence RNNs for Text Summarization”, The SIGNLL Conference on Computational Natural Language Learning (CoNLL), (2016), pp. 1-4.
- [6] P. Li, W. Lam, L. Bing and Z. Wang, “Deep Recurrent Generation Decoder for Abstractive Text Summarization”, In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)., (2017), pp. 2081-2090.
- [7] M. Yousefi-Azar and L. Hamey, “Text Summarization using Unsupervised Deep Learning”, Expert Systems with Applications, vol. 68, (2017), pp. 93-105.
- [8] K. Choi and C. Lee, “End-to-end Document Summarization Using Copy Mechanism and Input Feeding”, Journal of KIISE., vol. 44, no. 5, (2017), pp. 503-509.
- [9] Y. Kim, Y. Jernite, D. Sontag and A. M. Rush, “Character-Aware Neural Language Models”, In AAAI, (2016), pp. 2741-2749.
- [10] Y. Hayashi and H. Yanagimoto, “Generation with Recurrent Neural Network”, In 5th Advanced Applied Informatics (IIAI-AAI) International Congress., (2016), pp. 250-255.
- [11] R. Mihalcea and P. Tarau, “TextRank: Bringing Order into Texts”, In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, (2004), pp. 1-8.
- [12] W. Wang, D. Zhao and D. Wang, “Chinese News Event 5W1H Elements Extraction using Semantic Role Labeling”, Third International Symposium on Information Pressing., (2010), pp. 484-489.
- [13] S. Sharma, R. Kumar, P. Bhadana and S. Gupta, “News Event Extraction Using 5W1H Approach & Its Analysis”, International Journal of Science & Engineering Research., vol. 4, no. 5, (2013), pp. 2064-2067.
- [14] J. Yoon, Y. Chung and J. Lee, “Automatic Extractive Summarization of Newspaper Articles using Activation Degree of 5W1H”, Journal of KISS: Software and Applications. vol. 31, no. 4, (2004), pp. 505-515.
- [15] The National Institute of the Korean Language, “The Literal Style of Korean Newspaper Articles”, (1997).
- [16] CRF++. Software. <https://taku910.github.io/crfpp/>.

Authors



Meiying Ren
Ph.D Student in Daegu University.
Interested in Natural Language Processing.



Sinjae Kang
Professor, Daegu University.
Research Area: Natural Language Processing, Machine Learning