Nuclear Engineering and Technology 55 (2023) 421–431

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Retrieval methodology for similar NPP LCO cases based on domain specific NLP

No Kyu Seong [a, b], Jae Hee Lee [b], Jong Beom Lee [b], Poong Hyun Seong [a, *]

[a] Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, 291, Gwahak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea

[b] KHNP Central Research Institute, 70, 1312 beon-gil, Yuseong-daero, Yuseong-gu, Daejeon, 34101, Republic of Korea

A R T I C L E I N F O

Article history:
Received 17 June 2022
Received in revised form
3 September 2022
Accepted 26 September 2022
Available online 13 October 2022

Keywords:
Technical Specifications
Limiting Conditions for Operation
TF-IDF
Similarity
Information Retrieval
NLP

A B S T R A C T

Nuclear power plants (NPPs) have technical specifications (Tech Specs) to ensure that the equipment and key operating parameters necessary for the safe operation of the power plant are maintained within limiting conditions for operation (LCO) determined by a safety analysis. The LCO of Tech Specs that identify the lowest functional capability of equipment required for safe operation for a facility must be complied for the safe operation of NPP.

There have been previous studies to aid in compliance with LCO relevant to rule-based expert systems; however, there is an obvious limit to expert systems for implementing the rules for many situations related to LCO. Therefore, in this study, we present a retrieval methodology for similar LCO cases in determining whether LCO is met or not met. To reflect the natural language processing of NPP features, a domain dictionary was built, and the optimal term frequency-inverse document frequency variant was selected. The retrieval performance was improved by adding a Boolean retrieval model based on terms related to the LCO in addition to the vector space model. The developed domain dictionary and retrieval methodology are expected to be exceedingly useful in determining whether LCO is met.

1. Introduction

Nuclear power plants (NPPs) have technical specifications (Tech Specs) to ensure that the equipment and key operating parameters necessary for the safe operation of the power plant are maintained within limiting conditions for operation (LCO) determined by a safety analysis. The LCO of the Tech Specs that identify the lowest functional capability of equipment required for safe operation for the facility must be monitored, and if LCO is not met, the operators are required to take actions related to the LCO within completion time [1,2]. To comply with LCO, the operators prevent a situation that LCO is not met in advance through the surveillance test etc. and immediately detect LCO abnormality, and quickly determine whether the LCO is met or not met. If the LCO is not met, the operators declare the entry into the Actions of LCO and take the required actions within completion time, and Lastly, they document a series of procedures related to the LCO. If the above-mentioned procedures are not carried out properly, then they are not in compliance with LCO. Cases related to non-compliance with LCO occasionally occur in NPPs [3] and there have been many studies to prevent them.

Ragheb and Abdelhai developed the model-based analysis system for tracking of LCO and surveillance requirements [4] and Ragheb et al. created a production-rule analysis system having with a natural language interface [5]. Lidsky et al. studied computerized Tech Specs using the prolog [6]. Schlegelmilch and Jefferis introduced a system that enables utilities to better manage their nuclear plant operations by automating the LCO compliance decisions using an expert system [7]. Fiedler et al. developed TechSPEX which is a knowledge based advisory system for checking if an NPP complies with safety limits and the LCO [8]. Paiva and Schirru developed an expert system to apply, in real time, the rules contained in the LCO using Python language [9]. Schechter et al. developed an expert system and applied it at the Angra 2 NPP to assist in the resolution of LCO [10]. Oluwasegun and Jung explained that the Tech Specs for the control rod drive system and information regarding the LCO can

* Corresponding author. Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, 291, Gwahak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea.
E-mail address: phseong@kaist.ac.kr (P.H. Seong).

https://doi.org/10.1016/j.net.2022.09.028

be incorporated into a digital twin design in the form of an expert system [11]. Corrales et al. introduced an Alarm Processing and Diagnostic System (APDS) which facilitates alarm monitoring in a list format by applying alarm reduction techniques based on filters that adjust the length of the alarm list depending on the number of alarms and their priorities [12]. Kim and Jeong developed a Technical Specification Monitoring System (TSMS) which continuously monitors NPP operating status by using real-time operating parameters and alerts operators when the LCOs applicability is required [13]. Lee and Kim studied the design of a computerized operator support system for technical specification monitoring [14].

These studies can be broadly classified into those focusing on determination of whether LCO is met and those related to the detection of LCO abnormality. Regarding the studies for an expert system, there are obvious limits in implementing the rules for many situations related to LCO. Each situation related to LCO is different from case to case. Consequently, it is not possible to cover all the assumptions on which every decision is based; therefore, the decision suggested by experts can differ from the one made by the expert system. The expert system can be useful in solving distinct situations in specific equipment or systems. But it cannot reflect the domain knowledge that considers every circumstance of a complex nuclear power plant with many types of equipment and many systems. As for the studies focused LCO abnormality detection, detection is not a critical issue because there are other methods for recognizing LCO abnormalities, such as NPP alarm systems. The operators have more difficulty in determining whether LCO is met or not considering the various situations of NPP [15].

Therefore, in this study, we present a retrieval methodology for similar LCO cases in determining whether LCO is met or not. If similar LCO cases based on the operator's query are provided, it can be particularly useful to determine if LCO is met. For example, if a user retrieves data for the 'main steam isolation valve', and if the entry cases into the Actions of LCO related to the query are provided then the operator can quickly determine if LCO is met. To reflect the natural language processing (NLP) of NPP features, we built the nuclear domain dictionary and also selected the optimal Term Frequency-Inverse Document Frequency (TF-IDF) variant suitable for LCO retrieval. And Boolean retrieval model (BRM) based on terms related to LCO in addition to the Vector space model (VSM) was developed. Finally, the coverage of the built dictionary was validated by operators' queries constructed through a survey and case studies was conducted to evaluate the retrieval performance of the developed retrieval methodology. This paper is the first study how support operators make decisions by using text data related to LCO. Using the proposed retrieval methodology, it is expected that violation of LCO can be prevented in advance by assisting with the operator's quick determination by providing similar NPP LCO cases if an abnormal situation related to LCO occurs.

2. Analysis of LCO of Tech Specs at NPP

2.1. LCO in the safety margins

As defined in the standard Tech Specs of nuclear regulatory commission (NRC), LCO is the minimum requirements for ensuring safe operation of the unit, and the range of LCO in the safety margin is shown in Fig. 1 [16]. As described in Fig. 1, LCO is the smallest range of the safety margins, so the operator must periodically check whether the current plant state is within this range. If it is determined that the LCO is not met, the operator must take the required actions, such as power reduction and plant shutdown, to keep the plant in a safe state. If the operator evaluates the LCO too conservatively due to concerns about violating of the Tech Specs, the power plant utilization rate may decrease due to the frequent execution of required actions. On the other hand, if the operator too broadly evaluates LCO, the likelihood of an LCO violation may increase.

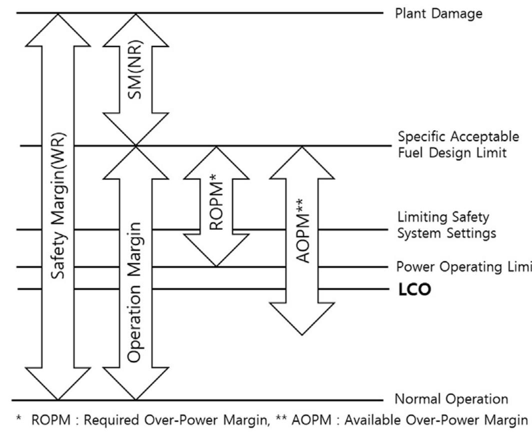

Fig. 1. Definition of safety margins in terms of thermal variables.

2.2. LCO structure

The Improved standard Tech Specs (ITS) in tabular format in 1992 consists of *LCO, applicability, actions, and surveillance requirements* [2]. The LCO to be monitored is determined according to its *applicability*, and *applicability,* which is based on the operation modes presented in Table 1, includes exceptional and additional conditions. *Actions* describes the required action to be taken when the LCO is not met, and *surveillance requirements* describes the inspection frequency and requirements to determine if LCO is periodically met. Fig. 2 shows the example of LCO 3.7.2 Main Steam Isolation Valve of the standard Tech Specs. The number of LCO, LCO with additional conditions of applicability and LCO including operability are tabulated in Table 2. If the parameters to be monitored are subdivided, such as in LCO '3.3.1 Reactor Protective System (RPS) Instrumentation - Operating', the total number of LCO increases. There are also many additional and exceptional conditions for *applicability* based on the plant mode, therefore, it is difficult for the operator to determine whether LCO is met or not met considering all situations related to LCO.

2.3. Compliance with LCO

Compliance with LCO requires performing the six procedures shown in Fig. 3. First, the situation in which LCO is not met should be prevented in advance if possible. The periodic surveillance test can be included in this procedure. Second, LCO abnormality must be detected immediately. Next after an LCO abnormality is detected, the operator determines whether LCO is met or not. If LCO is not met, then the shift reactor operator declares an entry into the Actions of LCO. After declaration, the required actions must be taken within completion time, and a series of results related to the LCO must be documented. Among the six procedures of the compliance with LCO, the rapid determination of whether LCO is met or not is of utmost importance because the completion time of Actions of LCO is dependent on the time of detection not on determination. Fig. 4 shows the allowable time to take required actions according to time of detection and time of determination [17].

Table 1
Operation modes of standard Tech Specs.

MODE	Title	Reactivity Condition (K_{eff})	% Rated Thermal Power	Average Reactor Coolant Temperature (°F)
1	Power Operation	≥ 0.99	> 5	NA
2	Startup	≥ 0.99	≤ 5	NA
3	Hot Standby	< 0.99	NA	≥ [350]
4	How Shutdown	< 0.99	NA	[350] > T_{avg} > [200]
5	Cold Shutdown	< 0.99	NA	≤ [200]
6	Refueling	NA	NA	NA

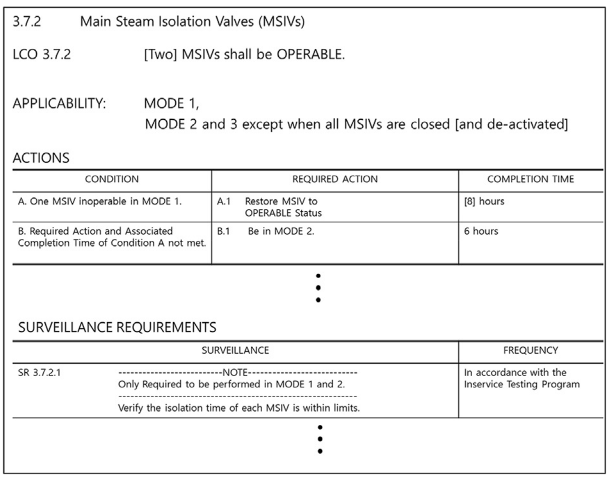

Fig. 2. LCO structure (LCO 3.7.2 Main Steam Isolation Valves of standard Tech Specs).

Table 2
LCO features of standard Tech Specs.

LCO No	Systems	NO. of LCO	NO. of LCO with additional conditions of applicability	NO. of LCO with Operability
3.1	Reactivity Control System	9	4	2
3.2	Power Distribution Limits	5	5	1
3.3	Instrumentation	13	8	13
3.4	Reactor Coolant System	18	8	10
3.5	Emergency Core Cooling Systems	5	3	4
3.6	Containment Systems	12	0	10
3.7	Plant Systems	19	10	15
3.8	Electrical Power Systems	10	6	9
3.9	Refueling Operations	6	5	4
	Total	**97**	**49**	**68**

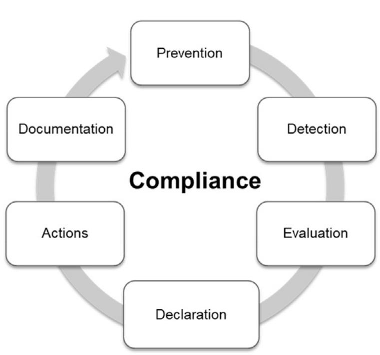

Fig. 3. Compliance with LCO

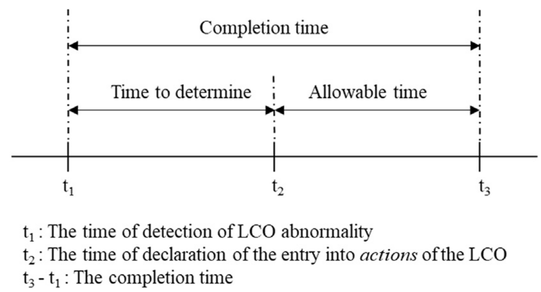

Fig. 4. The allowable time to take required actions.

2.4. Difficulty of compliance with LCO

As mentioned in Section 2.2, there are many LCO that must be monitored, as well as types of monitoring parameters [14]. Although the operation modes are classified into six modes, the applicability of LCO has additional conditions in addition to the operation modes such as LCO 3.1. 7 Mode 1 > 20% RTP. Accordingly, it is necessary to notify operators of an LCO abnormality. In Korean NPPs, LCO relevance is indicated on the alarm tile as shown in Fig. 5, so that the operator can recognize an LCO abnormality. As there is a path that can detect LCO abnormality, a support function for determination is needed to maximize the allowable time measures, as described in Section 2.3. As described in Table 2, in the case of an LCO that does not include operability, if an LCO abnormality is detected, then it is relatively easy to determine whether LCO is met or not. On the other hand, if an LCO includes operability, it is difficult to determine whether LCO is met because the operator's determination is additionally needed. However, the determination of operability is a difficult and high burden task. The simplified block diagrams for determining the operability of LCO 3.7.2 'MSIVs shall be operable' is depicted schematically in Fig. 6. In case, if the N_2 pressure is less than xx kg/cm^2, the specified safety function of MSIV cannot be performed; therefore, there is no difficulty in declaring the entry into the Actions of the LCO. However, in case, due to the redundancy of other trains to actuate the MSIV, the entry into the Actions of the LCO according to the operator's point of view can be determined differently. To solve the difficulty, providing similar LCO cases can make it significantly easier to determine if LCO is met. For example, if the similar LCO cases related to MSIV solenoid valve(A1) are provided, then the operator can quickly determine that LCO is not met.

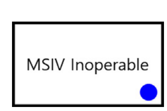

Fig. 5. An example of alarm tile with the mark related to LCO.

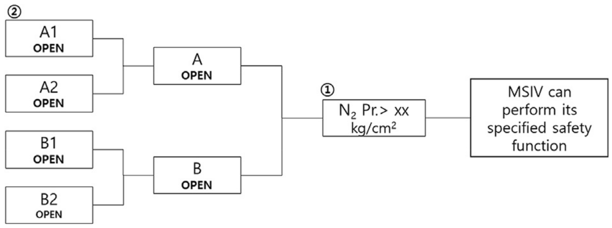

Fig. 6. The simplified block diagrams for determining operability of LCO 3.7.2.

3. Development of a domain specific dictionary for NPPs

An example of information retrieval is shown in Fig. 7. With respect to the user's query, similarity is obtained from the document collection through the retrieval module, and the ranking is determined according to the similarity between the query and the documents. Finally, a set of relevant documents is provided to the user.

3.1. Problems

Natural language processing refers to the ability of a computer to understand human language [18]. For a computer to understand natural language, the language must be converted into numbers that the computer can understand. Therefore, it is necessary to tokenize a sentence or word into a meaningful term and convert it into a number. It is also necessary to reflect the domain specific characteristics to tokenize a sentence or word into a meaningful term. If the NPP features are not reflected on the NLP, it may have a negative effect on the retrieval results. Representative types of problems are described, as follows.

1) Common nouns with dependent morpheme

If the user query and document are tokenized by whitespace, the set of retrieved relevant documents can be different even though the queries have the same meaning. For 'Jujeung-gigyeoglibaelbeu'(main steam isolation valve), 'Jueung-gigyeoglibaelbeu-ui' (of main steam isolation valve), the system recognizes the two queries as a completely different term and provides different results. Table 3 shows the similarity results of Query 1-A and Query 1-B for Problem 1). The LCO cases (Index 14135, 10561) not related to LCO 3.7.2 MSIV were retrieved. To overcome Problem 1), tokenization is required for independent and dependent morphemes through a morphological analyzer, and the morphological analyzer which is explained in Section 3.2.

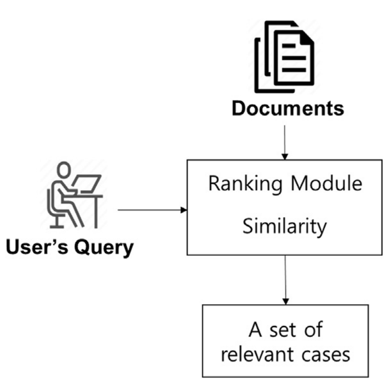

Fig. 7. Information retrieval process.

2) Synonym

The second problem is the processing of synonyms. In NPPs, abbreviations are used to express a large amount of information in a limited space. The abbreviations used in the main control room are included in the Human Factors Engineering Guideline (HFEG). While texts in design documents and drawings for equipment and systems are written in English, procedures and Tech Specs are written in Korean, so Korean texts and English texts are mixed. Therefore, synonym processing is important. Table 4 shows the similarity results of Query 2-A and Query 2-B for Problem 2). The completely different results are retrieved as synonyms are not processed.

3) Compound nouns separated by whitespace

The final problem is the processing of compound nouns separated by whitespace. The similarity results for the two queries are described in Table 5. The system recognizes the two queries differently so that the set of relevant results are different. The whitespace processing is important because the meaning of the text written in Korean can be changed depending on the whitespace between terms compared to the text written in English [18]. In NPPs, most terms are composed of compound nouns, and the operators can use compound nouns according to whitespace differently; therefore, it is necessary to build a compound nouns dictionary to recognize the compound nouns as a single term.

3.2. Morphological analysis

A morphological analysis is required to tokenize the dependent and independent morpheme and to identify the syntactic role of terms in a sentence. The morpheme is a minimum unit with meaning. If a term is subdivided into more than a morpheme, it loses its meaning [18]. To select a suitable Korean morphological analyzer, sentences were examined using the analyzers provided by the Python package konlpy and the Kakao morphological analyzer [19,20]. The tokenized results of two terms using each naive tokenizer where the domain dictionary is not added to are shown in Table 6. The expected results of the term 'Jagdong-yugudongpeompeu'(starting oil driven pump) by subdividing a term into three terms: 'Jagdong-yu'(starting oil) + 'Gudong'(driven) + 'Peompeu'(pump). However, each morphological analyzer analyzes terms differently. For example, in the case of the Hannanum analyzer, 'Jagdong-yugudongpeompeu'(starting oil driven pump) was recognized as one term while other morphological analyzers analyzed the term 'Jagdong-yu' differently, such as 'Jag' + 'dong-yu', or 'Jagdong' + 'yugu'. As the same result of term 2, 'Jujeung-gigyeoglibaelbeu' (main steam isolation valve) was not analyzed properly into 'Jujeung-gi'(main steam), 'Gyeogli'(isolation), and 'baelbeu'(valve). Therefore, if a morphological analyzer is used without the domain dictionary, the similarity results may be worse than if a morphological analyzer is not used. In this study, the morphological analysis was performed by adding the domain dictionary to Open Korean Text (OKT).

Table 3
An example of Problem 1).

(a) Query 1-A 'Jujeung-gigyeoglibaelbeu SollenoideuBaelbeu'			
Index	Title	Similarity	Relevance
6702	Jujeung-gigyeoglibaelbeu(V152) unjeonbulneung T/S jeog-yong, (MSIV(V152) inoperable T/S applied)	0.434	True
10213	Jujeung-gigyeoglibaelbeu(V014) unjeonbulganeung, (MSIV(V014) inoperable)	0.239	True
2971	Jujeung-gigyeoglibaelbeu(V151) unjeonbulneung(LCO 3.7.2), (MSIV(151) inoperable(LCO 3.7.2))	0.210	True
(b) Query 1-B 'Jueung-gigyeoglibaelbeu-ui SollenoideuBaelbeu'			
Index	Title	Similarity	Relevance
14135	[LCO 3.9.8] bojogeonmul jung-anghwangipaen jeongji, [LCO 3.9.8] (Aux. building main fan stopped)	0.171	False
3600	Jujeung-gichadanbaelbeu 'A' Inoperable (MSIV 'A' inoperable)	0.170	True
10561	LCO 3.4.16 wonjalonaeng-gagjaebaegigyetong (LCO 3.4.16 Reactor coolant gas vent system)	0.166	False

Table 4
An example of Problem 2).

(a) Query 2-A 'Jujeung-gigyeoglibaelbeu SollenoideuBaelbeu'			
Index	Title	Similarity	Relevance
6702	Jujeung-gigyeoglibaelbeu(V152) unjeonbulneung T/S jeog-yong (MSIV(V152) inoperable T/S applied)	0.434	True
10213	Jujeung-gigyeoglibaelbeu(V014) unjeonbulganeung (MSIV(V014) inoperable)	0.239	True
2971	Jujeung-gigyeoglibaelbeu(V151) unjeonbulneung(LCO 3.7.2) (MSIV(151) inoperable(LCO 3.7.2))	0.210	True
(b) Query 2-B 'MSIV SollenoideuBaelbeu'			
Index	Title	Similarity	Relevance
10417	MSIV 1 gae unjeonbulganeung (one of MSIVs inoperable)	0.313	True
4205	MSIV-151 jagdong-yu nuyu sudong chadan (MSIV-151 starting oil leakage manual shutoff)	0.277	False
10617	Y1(LCO) MSIV 'B' unjeonbulneung (Y1(LCO) MSIV 'B' Inoperable)	0.252	True

Table 5
An example of Problem 3).

(a) Query 3-A 'Jujeung-gigyeoglibaelbeu SollenoideuBaelbeu'			
Index	Title	Similarity	Relevance
6702	Jujeung-gigyeoglibaelbeu(V152) unjeonbulneung T/S jeog-yong (MSIV(V152) inoperable T/S applied)	0.434	True
10213	Jujeung-gigyeoglibaelbeu(V014) unjeonbulganeung (MSIV(V014) inoperable)	0.239	True
2971	Jujeung-gigyeoglibaelbeu(V151) unjeonbulneung(LCO 3.7.2) (MSIV(151) inoperable(LCO 3.7.2))	0.210	True
(b) Query 3-B 'Jujeung-gi gyeogli baelbeu SollenoideuBaelbeu'			
Index	Title	Similarity	Relevance
4204	Jujeung-gi chadanbaelbeu(MSIV-151) dongjag bulneung (MSIV(MSIV-151) inoperable)	0.268	True
9056	AB-HV108(MSIV 'A') FBM kadeu gyoche (AB-HV108(MSIV 'A') FBM card replacement)	0.267	False
4400	Jujeung-gi yang-ion jeondodo gamsi bulganeung (Main steam cation conduntivity monitoring inoperable)	0.262	False

Table 6
Tokenized results for terms.

	Tokenized results of term 1	Tokenized results of term 2
Hannanum	(Jagdongyugudongpeompeu)	(Jujeung-gigyeoglibaelbeu)
Kkma	(Jagdong) (yu) (gudong) (peompeu)	(Ju) (jeunggi) (gyeogli) (baelbeu)
Komoran	(Jagdong) (yugu) (dong) (peompeu)	(Ju) (jeunggi) (gyeogli) (baelbeu)
OKT	(Jag) (dong) (yugu) (dong) (peompeu)	(Ju) (jeunggi) (gyeogli) (baelbeu)
Mecab	(Jag) (dongyu) (gudong) (peompeu)	(Ju) (jeunggi) (gyeogli) (baelbeu)
Khaiii	(Jagdong) (yu) (gudong) (peompeu)	(Jujeung) (gigyeogli) (baelbeu)

Term 1: 'Jagdong-yugudongpeompeu'(starting oil driven pump) - Expected results: Jagdong-yu(starting oil) + Gudong(driven) + Peompeu(pump).
Term 2: 'Jujeunggigyeoglibaelbeu'(main steam isolation valve) - Expected results: Jujeunggi(main steam)+Gyeogli(isolation) + Baelbeu(valve).

3.3. The data to build a domain dictionary

1) Glossary of nuclear power

There is a total of 9221 terms in the nuclear power glossary and examples of the Korean-English pairs are shown in Table 7 [21]. Because more than 85% of terms consist of compound nouns, it is necessary to separate them into common nouns to acquire more common nouns. For example, a compound noun such as 'Bisang-dijelbaljeongi'(emergency diesel generator) can be distinguished by three common nouns: 'Bisang'(emergency), 'Dijel' (disel) and 'Baljeongi'(generator). By extracting 3587 additional common nouns from the compound nouns, 4864 common nouns and 7994 compound nouns were acquired.

2) Failure modes by the equipment

The failure modes by the equipment must be added to the domain dictionary because the entry into the Actions of the LCO is mainly declared due to the occurrence of equipment failure. Therefore, when an LCO-related equipment failure occurs, the operator records the equipment failure modes as the cause of the LCO; therefore, it is necessary to develop a related dictionary to recognize the failure modes by the equipment. For example, in the case of a diesel generator, failure types include *start failure*, *low output*, *stop failure*, *malfunction start*, *external leakage* and so on [22]. 3 common nouns and 66 compound nouns were added to the domain dictionary.

3) Abbreviations

Abbreviations have been used in NPPs to provide more necessary information in the limited display space and to communicate effectively in the main control room [23]. HFEG defines a total of 1443 abbreviations used in the main control room, such as level (lvl), pump (pp), RCS cold leg temperature (Tcold), etc. 780 compound nouns and 665 common nouns were added to the domain dictionary.

4) Stopword

A stopword dictionary consists of words that are not related or words that have relatively low importance. The types of words added to the stopword dictionary can be classified into four types.

- Terms automatically included when downloading the LCO cases from the database.
- Terms related to Tech Specs
- Terms with low relevance in determining similar LCO cases
- Terms related to numbers

Table 7
Examples of nuclear power glossary.

No	Korean	English
16	Heubsu	Absorption
…	...	...
48	Gasoggi	Accelerator
…	...	...
2633	Bisangdijelbaljeongigyetong	Emergency Diesel Generator System (EDGS)
...	...	...

With respect to terms related to numbers, the two sentences below can be tokenized as follows.

Sentence 1: EDG 01A starting air leakage (EDG, 01, A, starting, air, leakage)

Sentence 2: RCP 01A seal flow hunting (RCP, 01, A, seal, flow, hunting)

Because token '01' exists in both sentences after morphological analysis, unrelated LCO cases can be retrieved using some queries. All numbers, therefore, were excluded to improve the retrieval performance.

Problems 1), 2) and 3) can be solved by a domain specific dictionary with a total of 5530 common nouns and 8840 compound nouns. Examples of the constructed dictionary are shown in Tables 8 and 9.

4. Development of a domain specific retrieval methodology

Fig. 8 describes the procedure for the proposed retrieval methodology.

1. The entry cases into Actions of the LCO and an operator's query are tokenized by the tokenizer with a domain dictionary.

   For example,

   - Original Case: DPPS coincidence logic ch. trouble
   - Tokenized Case: DPPS coincidencelogic channel trouble.

   The terms 'coincidence' and 'logic' are merged due to the compound nouns dictionary.
   The term 'ch.' is converted to 'channel' due to the common nouns dictionary.

2. Each divided token is compared with the terms related to LCO included in the BRM. If it is included in terms related to LCO, only LCO cases that contain the token are extracted; otherwise, all LCO case are used.

   For example, DPPS [True], conincedencelogic [True], channel [False], trouble [False]. The set of LCO cases containing 'DPPS' or 'conincedencelogic' are retrieved using BRM.

3. The operator's query and the set of LCO cases (or all the LCO cases) are vectorized by the TF-IDF weighting scheme.
4. The similarity is calculated between the query and each case of the set of LCO cases (or all the LCO cases). The similarity between 0 and 1 can be obtained through the cosine similarity. The closer the result is to 1, the higher the similarity between the user's query and the document is determined.

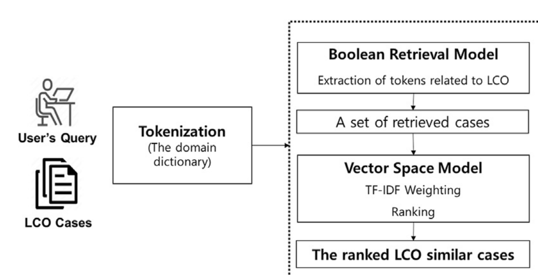

Fig. 8. The framework of retrieval methodology for LCO similar cases.

Table 8
Examples of the common nouns dictionary.

Common nouns	Synonym1	Synonym2	Synonym3	Synonym4	Synonym5
Gaabgi	Pressurizer	PZR			
Gyetong	System	sys	..	..	
Naeng-gagjae	Coolant	CLNT			
Bulneung	Inoperable	Inop	Disabled	DIS	
Baelbeu	Valve	VLV	..	..	
Suwi	Level	LVL			
Jeeogi	Controller	Jojeolgi	Regulator	CTRL	REGR
Peompeu	Pump	PMP	PP	..	
…					

Table 9
Examples of the compound nouns dictionary.

Compound nouns	Noun1	Noun2	Noun3	Noun4	…
Nosimbohoyeonsangi, CPC	Nosim	boho	yeonsangi		
Jujeunggigyetong, MS	Jujeunggi	gyetong			
Jujeunggigyeoglibaelbeu, MSIV	Jujeunggi	gyeogli	baelbeu		
Teobinbaljeongigamsijeeogyetong, TM	Teobin	baljeongi	gamsi	jeeo	gyetong
Teobinbaljeongigyecheugjeeogyetong, TN	Teobin	baljeongi	gyecheug	jeeo	gyetong
Chugbanghyangchullyeogpyeoncha, ASI	Chugbanghyang	chullyeog	pyeoncha		
Chullyeog-unjeonjehanchi, POL	Chullyeog	unjeon	jehanchi		
…					

5. The top 10 results are shown in the order of similarity.

4.1. Vector space model (VSM)

The representation of a set of documents as vectors in a common vector space is known as the vector space model and is fundamental to a host of information retrieval operations ranging from scoring documents on a query, document classification and document clustering [24,25].

$$\boldsymbol{d}_j = (w_{1,j}, w_{2,j}, \ldots, w_{t,j})$$

$$\boldsymbol{q} = (w_{1,q}, w_{2,q}, \ldots, w_{t,q})$$

$$\boldsymbol{d} = document, \boldsymbol{q} = query, j = document\ number, w = word$$

If a term exists in the document, its value in the vector can be weighted non-zero value. If a document and a query can be expressed as each vector, the cosine similarity between the document and the query can be obtained through the dot product [25].

$$similairty(\boldsymbol{d}_j, \boldsymbol{q}) = \frac{\boldsymbol{d_j} \bullet \boldsymbol{q}}{\|\boldsymbol{dj}\|\|\boldsymbol{q}\|}$$

$$\boldsymbol{d} = document, \boldsymbol{q} = query, j = document\ number$$

A calculation result between 0 and 1 can be obtained through the cosine similarity, and the closer the result is to 1, the higher the similarity between the user's query and the document is determined. Therefore, using this result, it is possible to obtain a rank for LCO cases similar to a query. TF-IDF is used to convert the tokenized words into numbers, which will be discussed in detail in the following section.

4.2. Term frequency – inverse document frequency (TF-IDF)

The TF-IDF, suggested by Salton, has been widely used in the information retrieval field due to its simplicity and effectiveness [26–28].

- Term frequency (tf): The number of times that the term t occurs in the document.
- Inverse document frequency (idf): The inverse of the number of documents in which the term occurs.

$$tf - idf_{t,d} = tf_{t,d} \times idf_t$$

$$t = term, d = document$$

For example, if 1-A of Table 10 is applied to the terms 'was' and ‘system’ included in the following LCO case.

‘Digital plant protection system was failed because its power system was inoperable’

The terms 'was' and ‘system’ occur twice, and all other terms occur once. Assuming there are 90 cases containing the term 'was' and 30 cases containing ‘system’ in a total of 100 LCO cases, the terms ‘was’ and ‘system’ are converted into 0.092 and 1.046, respectively.

TF-IDF variants are shown in Table 10 [24]. In this study, the optimal TF-IDF variant suitable for similar LCO case retrieval was

Table 10
TF-IDF variants.

	Term Frequency (TF)		Inverse Document Frequency (IDF)
1	$tf_{t,d}$	**A**	1
2	$\begin{Bmatrix} 1 + \log(tf_{t,d}) & if\ tf_{t,d} > 0 \\ 0 & otherwise \end{Bmatrix}$	**B**	$\log \frac{N}{df_t}$
3	$a + \frac{(1-a) \times tf_{t,d}}{max_t(tf_{t,d})} \quad a = [0, 1]$ *If, a = 0, 0.4, 0.7*	**C**	$\max\left\{0, \log \frac{N - df_t}{df_t}\right\}$
4	$\begin{Bmatrix} 1 & if\ tf_{t,d} > 0 \\ 0 & otherwise \end{Bmatrix}$		
5	$\frac{1 + \log(tf_{t,d})}{1 + \log(ave_{t \in d}(tf_{t,d}))}$		

selected through the evaluation method described in the following section.

4.3. Mean average precision (MAP)

The mean average precision was used to select the optimal TF-IDF variant. MAP has been shown to have especially good discrimination and stability [24]. The precision and the average precision were used to calculate MAP.

- Precision: The fraction of the retrieved documents that are relevant

$$Precision = P(relevant|retrieved) = \frac{The\ number\ of\ relevant\ items\ retrieved}{The\ number\ of\ retrieved\ items}$$

- Average precision (AP): The average of the precision value obtained for the set of top k documents existing after each relevant document is retrieved.

$$Average\ Precision = \frac{1}{m_j}\sum_{k=1}^{m_j} Precision\left(R_{jk}\right)$$

$$m = The\ number\ of\ retrived\ documnets$$
$$R_{jk} = The\ set\ of\ ranked\ retrieval\ results$$

- Mean average precision (MAP): The mean of the AP over information needs.

$$Mean\ Average\ Precision = \frac{1}{|\mathbf{Q}|}\sum_{j=1}^{|\mathbf{Q}|}\frac{1}{m_j}\sum_{k=1}^{m_j} Precision\left(R_{jk}\right)$$

$$\mathbf{Q} = Quires,\ m = The\ number\ of\ retrived\ documnets$$
$$R_{jk} = The\ set\ of\ ranked\ retrieval\ results$$

The AP for the representative queries acquired through the operators' survey are shown in Table 11 and the $tf_{t,d}$ and $\log\frac{N}{df_t}$ with the highest MAP values were selected. Table 12 shows MAP for the representative quires.

4.4. Boolean retrieval model based on terms related LCO

TF-IDF, which is used as a factor for ranking in Internet search engines, only considers term frequency and inverse document frequency without considering the importance of the term [29]. Consequently, even if an important word is not written in the document, the similarity between a query and a document can be calculated as high. To solve this problem, it is necessary to add BRM that can reflect the importance based on LCO-related words. The BRM is a model for information retrieval in which we can pose any query which is in the form of a Boolean expression of terms, that is, in which terms are combined with the operators AND, OR, and NOT [24]. If a problem occurs in power plant equipment, the cause of the problem may vary, but the symptoms that appear may be similar. For example, a pump or valve can cause ‘leakage’, ‘trouble’ and ‘inoperable’ symptoms. User queries mainly include symptoms for devices, and since symptoms can appear common to other devices, irrelevant LCO cases can be retrieved. For example, in the case of a ‘DPPS coincidence logic channel trouble’ query, the term 'DPPS' is directly related to the LCO while the term ‘channel trouble’ is not

Table 11
Average precision for representative queries.

	Type	TF			1+log(TF)			Augmented(a = 0)			Bool			Log_ave		
		A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
Query 1																
1	NT/VSM	0.517	0.436	0.436	0.486	0.302	0.302	0.436	0.227	0.227	0.336	0.227	0.227	0.336	0.252	0.252
2	NTD/VSM	0.604	0.616	0.616	0.57	0.629	0.629	0.402	0.486	0.486	0.496	0.496	0.496	0.467	0.55	0.55
Query 2																
1	NT/VSM	0.085	0.44	0.402	0	0.143	0.143	0	0	0	0	0.143	0.143	0	0.085	0.085
2	NT/VSM	0.399	0.64	0.64	0	0.212	0.212	0	0	0	0	0.177	0.177	0	0.085	0.085
Query 3																
1	NT/VSM	0	0.153	0.12	0	0	0	0	0.143	0.143	0	0	0	0	0	0
2	NTD/VSM	0.758	0.751	0.72	0.566	0.627	0.602	0.327	0.387	0.411	0.541	0.536	0.549	0.197	0.269	0.269
Query 4																
1	NT/VSM	0.034	0.065	0.065	0.119	0.164	0.164	0.075	0.086	0.069	0.146	0.191	0.191	0.103	0.119	0.119
2	NTD/VSM	0.419	0.469	0.469	0.227	0.227	0.227	0.01	0.01	0.01	0.191	0.191	0.191	0.164	0.164	0.164

Naïve Tokenizer (NT), Naïve Tokenizer with the Dictionary (NTD), Vector Space Model (VSM).
A: 1, B: $\log\frac{N}{df_t}$, C: $\max\left\{0, \log\frac{N-df_t}{df_t}\right\}$

Table 12
Mean average precision.

	TF			1+log(TF)			Augmented(a = 0)			Bool			Log_ave		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
MAP	0.352	**0.446**	0.434	0.246	0.288	0.285	0.156	0.167	0.168	0.214	0.245	0.247	0.158	0.191	0.191

A: 1, B: $\log\frac{N}{df_t}$, C: $\max\left\{0, \log\frac{N-df_t}{df_t}\right\}$

Table 13
Similarity results related to the query.

Index	Title	Similarity	Relevance
5959	DPPS Ch 'D' dongsinonli peuloseseo gojang (DPPS Ch 'D' coincidence logic processor fail)	0.214	True
872	Class 1E Ch A inbeoteo Trouble (Class 1E Ch A inverter trouble)	0.207	False
6401	DPPS A chaeneol uhoe (DPPS A channel bypass)	0.198	True
1517	Class 1E Ch B inbeoteo Trouble (Class 1E Ch B inverter trouble)	0.181	False

Table 14
Terms related to LCO.

LCO	Korean	English
…	…	…
3.3.1	Wonjalobohogyetong	RPS, DPPS
3.3.4	Gonghagjeog-anjeonseolbijagdong-gyetong	ESFAS, DPPS
3.4.1	Wonjalonaeng-gagjaegyetong	RCS
3.7.2	Jujeung-gigyeoglibaelbeu	MSIV
3.8.1	Bisangdijelbaljeongi	EDG
…	…	…

directly related to the LCO. Accordingly, the weight of the 'channel trouble' must be different from the weight of 'DPPS' because 'channel trouble' is a symptom that occurs in many other equipment. Table 13 shows the retrieval results before BRM is added to VSM. Indexes 872 and 1517, not related to DPPS, were retrieved. Therefore, if a term related to the LCO exists in the query, first the set of documents including the term is obtained, and the similarity is calculated using the TF-IDF in the obtained document set to improve the retrieval performance. A total of 110 Korean-English term pairs related to LCO were constructed, and an example is shown in Table 14.

5. Validation of the developed dictionary and retrieval methodology

5.1. The developed dictionary

A survey of twelve operators with a Shift Reactor Operator (SRO) or Rector Operator (RO) license was conducted in March 2022 to validate the developed dictionary and retrieval methodology. A total of 120 operators' queries were obtained through the survey. Four types of operators' queries are classified in Table 15.

Table 15
Representative types of operators' queries.

Representative types and examples
Type 1: Query based the plant parameters e.g. RCS jeoongwan ondo gamso (RCS cold leg decrease)
Type 2: Query based the plant alarms e.g. DPPS dongsinonli chaeneol Trouble (DPPS coincidence logic channel trouble)
Type 3: Query based the symptom of equipment e.g. EDG gidong-yong gong-gi nuseol (EDG starting air leakage)
Type 4: Query based the parts of equipment e.g. MSIV sollenoideu baelbeu (MSIV solenoid valve)

1) Korean

Table 16 shows the coverage analysis results for a total of 93 terms by extracting Korean terms from 120 queries. As the NPP glossary and failure modes are added, it can be observed that the coverage is widened from 77.78% to 97.8% compared to the only naive tokenizer. Compound nouns such as 'Sollenoideu'(solenoid) + 'Baelbeu'(valve) and 'Gong-geub'(supply) + 'Gigi'(device), written as 'Solbaelbeu'(solenoid valve) and 'Gong-geubgi'(supply device), were not tokenized properly; therefore, additional research that can reflect the various terms used by operators is required.

2) English

Coverage was validated by checking whether the tokenized English term exists in the dictionary because the naive tokenizer tokenized English terms by whitespace. The operators expressed the terms related to most of the devices and systems in queries with abbreviations in the case of English. It was observed that the coverage was improved by adding the abbreviations of HFEG. The coverage analysis is shown in Table 17. As in the Korean terms, expressing compound nouns as abbreviations (such as AFPP, aux feed water pump) were not properly tokenized. Also, the use of abbreviations P/P (pump) and V/V (valve), not included in the HFEG, was observed. These terms must be included in the abbreviations and managed.

Table 17
Coverage results for English term.

Glossary	Coverage (%)	Remarks
Naive tokenizer + NPP glossary	43/60 = 71.7	INOP, LVL, ESW, Hx, sys, Vlv, PP etc
Naive tokenizer + NPP glossary + Abbreviations of HFEG	52/60 = 86.7	AFPP, VV, P/P, V/V

5.2. The developed retrieval methodology

The proposed retrieval methodology was validated using a total of 14,227 the entry cases into the Actions of the LCO in Korean NPPs from 2003 to 2021. The evaluation method was as follows:

1. The representative queries were selected for each query type described in Table 15.

Table 16
Coverage results for Korean terms.

Glossary	Coverage (%)	Remarks	
		Expected	Results
Naive-tokenizer	70/93 = 77.78	Geubsu (feedwater)	(Geub), (su)
		Jisigye (indicator)	(Jisi), (gye)
Naive tokenizer + NPP glossary	85/93 = 91.4	Odongjag (malfunction)	(O), (dongjag)
		Gojindong (high vibration)	(Go), (jindong)
Naive tokenizer + NPP glossary + Failure mode code	91/93 = 97.8	Solbaelbeu (solenoid vavle)	(Sol), (baelbeu)
		Gonggeubgi (supply device)	(Gonggeub), (gi)

Table 18
Average precision of representative cases.

	Types	Case1	Case2	Case3	Case4
1	NT/VSM	0.436	0.44	0.153	0.065
2	NT/VSMB	**0.605**	**0.731**	**0.599**	**0.436**
3	NTD/VSM	0.616	0.64	0.751	0.469
4	NTD/VSMB	**0.738**	**0.817**	**0.905**	**0.969**

Naïve Tokenizer (NT), Naïve Tokenizer with the Dictionary (NTD).
Vector Space Model (VSM), Vector Space Model with Boolean retrieval Model (VSMB).

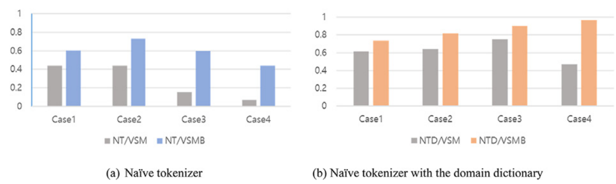

Fig. 9. Average precision with/without the BRM.

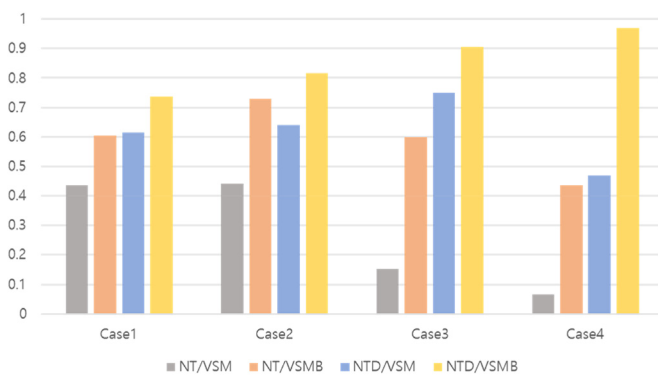

Fig. 10. Average precision for all cases.

2. Experts constructed LCO relevance for the selected queries among the entry cases into the Actions of the LCO.
3. For each representative query, the average precision was calculated for the set of ranked retrieval results from the top result until the 10th result.

To verify the effect of adding a developed dictionary, the results of the case with and without the dictionary in the naive tokenizer were compared. Next, to verify the effect of the developed retrieval methodology, the results of the case with and without the BRM in the naive VSM were compared. Table 18 shows the average precision according to the developed dictionary and retrieval methodology. The retrieval results are improved according to the addition of the BRM. There are many power sources, parts, root valves, and common properties that perform the same function in NPP; therefore, it is essential to construct a BRM that obtains related sets through the terms related to LCO.

Fig. 9 shows the retrieval performance for the developed retrieval methodology. For each case, the retrieval performance was increased by adding BRM based on terms relevant to LCO to the vector space model. The average precision for all cases is shown in Fig. 10. Each effect of retrieval performance according to the addition of the dictionary and according to the addition of the BRM was different in each case.

6. Conclusion

The LCO of Tech Specs that are the lowest functional capability of performance should be complied. However, determining whether LCO is met or not is an arduous task for operators. Therefore, this study was conducted to help determine whether LCO is met or not by retrieving similar LCO cases to operator's queries using the entry cases into the Actions of the LCO.

In order to use unstructured text data, operators' queries and entry cases into Actions of LCO were tokenized through a Korean morphological analyzer. A domain dictionary was developed to improve the tokenization results. A developed dictionary with a total of 5530 common nouns and 8840 compound nouns was constructed. Our dictionary was validated by covering Korean terms and English terms for operators queries through the survey.

The optimal TF-IDF variant suitable for similar LCO case retrieval was selected through the mean average precision evaluation method. To improve the retrieval performance, a retrieval methodology that adds a Boolean retrieval model based on terms related to LCO to the vector space model was suggested. The effect of the suggested retrieval methodology was demonstrated for the representative operators' quires. The developed domain dictionary and retrieval methodology are expected to be exceedingly useful in determining whether LCO is met or not met.

With respect to data quality, some LCO cases did not have records relevant to causes for LCO. To improve the retrieval performance, data quality is also important; therefore, a more accurate data need to be accumulated. In addition, it is expected that it would be helpful to operators if recommended cases considering the various plant situations and the characteristics of the operators and LCO are provided.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] https://www.nrc.gov/reading-rm/basic-ref/glossary/technical-specifications.html.
[2] U.S., Nuclear regulatory commission (NRC), standard technical specifications combustion engineering plants revision 4.0 volume 1, specifications NUREG-1432- U.S, Nucl. Regul. Commission (2012) 1.
[3] http://opis.kins.re.kr/.
[4] M. Ragheb, M. Abdelhai, Model-Based Analysis System for Tracking of Limiting Conditions for Operation and Surveillance Requirements in Power Plants. Applications of Artificial Intelligence V, International Society for Optics and Photonics, 1987, pp. 138–145.
[5] M. Ragheb, M. Abdelhai, J. Cook, Producton-rule analysis system for nuclear plant technical specifications tracking, in: Artificial Intelligence and Other Innovative Computer Applications in the Nuclear Industry, Springer, Boston, MA, 1988, pp. 673–680.
[6] L.M. Lidsky, et al., The use of prolog for computerized technical specifications, in: Artificial Intelligence and Other Innovative Computer Applications in the Nuclear Industry, Springer, Boston, MA, 1988, pp. 681–688.
[7] W.H. Schlegelmilch, R.P. Jefferis, Interface of an expert system with on-line data for technical specification monitoring, in: Artificial Intelligence and Other Innovative Computer Applications in the Nuclear Industry, Springer, Boston, MA, 1988, pp. 715–720.
[8] U. Fiedler, S. Schalm, K. Pranckeviciute, Knowledge Based Management of Technical Specifications in the Nuclear Industry, 1992, p. 131.
[9] G.V. Paiva, R. Schirru, Application of an Expert System for Real Time Diagnosis of the Limiting Conditions for Operation in Nuclear Power Plants, 2015.
[10] A. D. Schechter, A.S. Nicolau and R. Schirru, Expert System Applied at the Angra-2 Nuclear Power Plant to Assist in the Resolution of Limiting Conditions for Operation.

[11] A. Oluwasegun, J.C. Jung, The application of machine learning for the prognostics and health management of control element drive system, Nucl. Eng. Technol. 52 (10) (2020) 2262–2273.
[12] C. Corrales, A. Fuentes, F. Ortega, A New APDS Approach-Alarm Processing Based on Alarm Prioritization-201, vol. 555, American Nuclear Society-ANS, La Grange Park, IL 60526 (United States), 2017. North Kensington Avenue,.
[13] Y. Kim, J.J. Jeong, Development of a Technical Specification Monitoring System Using Real-Time Operating Parameters in OPR1000, Transactions of the Korean Nuclear Society Autumn Meeting, 2019.
[14] S. Lee, J. Kim, Lee, Design of computerized operator support system for technical specification monitoring, Ann. Nucl. Energy 165 (2022), 108661.
[15] S. Yu, N. Seong, Y. Kim, J. Lee, Conceptual Design of Operator Support System for Limiting Condition for Operation, Transactions of the Korean Nuclear Society Virtual Spring Meeting, 2020.
[16] S. Jang, An Approach for Justifying Safety Margin in the Risk-Informed Integrated Decisionmaking, Transactions of the Korean Nuclear Society Spring Meeting, 2019.
[17] N. Seong, J. Lee, J. Lee, Monitoring Methodology for Limiting Conditions for Operation of Technical Specification, Transactions of the Korean Nuclear Society Spring Meeting, 2022.
[18] H. Lim, Natural Language Processing Bible, second ed., Human Science, 2020.
[19] https://konlpy.org/en/latest/.
[20] https://github.com/kakao/khaiii.
[21] https://www.data.go.kr/.
[22] Korea Hydro Nuclear Power, Standard Maintenance Procedure, 2022.
[23] KEPCO Engineering & Construction, Human Factors Engineering Guideline, 9-750-J410-001, 2012.
[24] C.D. Manning, P. Raghavan, H. Schütze, Introduction to information retrieval, Nat. Lang. Eng. 16 (1) (2010) 100–103.
[25] https://en.wikipedia.org/wiki/Vector_space_model.
[26] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, Inf. Process. Manag. 24 (5) (1988) 513–523.
[27] J. Ramos, Using tf-idf to determine word relevance in document queries, in: Proceedings of the First Instructional Conference on Machine Learning, 2003, pp. 29–48.
[28] J. Son, S.B. Kim, H. Kim, S. Cho, Review and analysis of recommender systems, J. Kor. Instit. Indus. Eng. 41 (2) (2015) 185–208.
[29] K. Lee, Sentence Korean Embeddings Using Korean Corpus, Acorn, 2020.