

# Development of Nuclear Industry Information Gathering and Analysis System and Denied Persons Information Gathering System for Preventing Illegitimate Export of Trigger List Items in Korea

Sangjun Lee

*Korea Institute of Nuclear Nonproliferation and Control, 1418, Yuseong-daero, Yuseong-gu, Daejeon 34101, Republic of Korea*

(Received May 11, 2023 / Revised June 13, 2023 / Approved July 3, 2023)

---

In South Korea, the exporters of items related to nuclear power generation are diversified. Consequently, there is a risk of illegitimate export by companies failing to recognize the export control system because the awareness about this system for the strategic items among the subcontractors of nuclear power facilities is limited. To prevent illegitimate export of the strategic items, it is necessary to conduct outreach activities regarding the export control system for the related companies. Additionally, the exporters and export license examiners should consider whether an export target is on the Denial List, who may divert the strategic items to weapons of mass destruction. Therefore, the Korea Institute of Nuclear Nonproliferation and Control developed two systems for controlling illegitimate export of the Trigger List items. The first system, Nuclear Industry Information Collection and Analysis System, can gather information about the key nuclear industries in Korea and analyze the dealing of strategic items. The second system, Denied Persons Information Gathering System, can regularly gather information about the denied persons and provide the updated data to the exporters and regulatory examiners. These two systems can be used for outreach activities and export license examination to prevent illegitimate export of the strategic items.

**Keywords:** Illegitimate export, Strategic items, Information collection, Denial list, the Export control

---

\*Corresponding Author.

Sangjun Lee, Korea Institute of Nuclear Nonproliferation and Control, E-mail: leesj01406@kinac.re.kr, Tel: +82-42-860-9870

## ORCID

Sangjun Lee

<http://orcid.org/0000-0002-9841-5065>

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License [<http://creativecommons.org/licenses/by-nc/3.0>] which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

## 1. Introduction

Leakage of nuclear materials as well as relevant equipment and technologies should be controlled to prevent nuclear power from being used for military purposes, including weapons and terrorism. Strategic items refer to the items that require an export license by the governments, in pursuit of international peace, safety, and national security according to the principles of the international export control regimes. Representative examples of strategic items include the materials, software, and technologies that are used to manufacture and develop weapons of mass destruction and their means of transportation, such as missiles. South Korea designates and notifies the items that require limitations, such as an export license, according to the Foreign Trade Act and the principles of the multilateral export control regimes, and controls the export of the items according to the Foreign Trade Act [1]. In addition, the international community provides the Denial List to prevent diversion of the strategic items. The ‘denied persons’ refer to persons with whom the trade is limited or should be carried out with caution to seek international security and world peace. Whether a denied person is an export target, buyer or final cargo receiver is included in the criteria of the export license of the strategic items according to the Public Notice on Trade of Strategic Items [2].

However, despite export control of the strategic items, the number of illegitimate exports and unlicensed exports of strategic items uncovered by the intelligence agency of Korea has remained constant each year. In addition, due to changes in policies and environmental conditions regarding nuclear power generation, the target countries of export of trigger list items and the exporters of the items have been expanded [3]. In particular, small and medium-sized enterprises dealing with the goods, equipment, and technologies related to enrichment and reprocessing may be involved in the illegitimate export of strategic items due to the low recognition of the export control system for the strategic items. Therefore, the need for incorporating the Korean

companies dealing with the nuclear power related goods into the export control system has been raised to prevent illegitimate export of the strategic items. The present study was conducted to develop a method for satisfying this need, and the results from this study may be applied to the determination of new subjects of export control and the prevention of export to denied persons.

## 2. Overview of the Nuclear Industry Information Gathering and Analysis System

As described above, the Nuclear Industry Information Gathering System was developed to proactively identify the companies that are considered to be dealing with strategic items in order to prevent, in advance, the export of such materials without a license [4]. The Korea Institute of Nuclear Nonproliferation and Control conducts outreach activities about the export control system according to the list of companies dealing with the strategic items based on the cooperative companies of the Korea Hydro & Nuclear Power, the only operator of nuclear power plants in South Korea. However, the other companies dealing with nuclear strategic items that are not included in the list are excluded from the outreach activities. The simplest way of identifying more companies dealing with the strategic items is to employ a specialized search engine to find the companies dealing with each of the strategic items and determine whether they actually deal with the strategic item one by one. However, this approach is inefficient and ineffective, because information about the same companies is repeatedly searched, much of the information from general news articles and blogs is unrelated to the companies, and the method requires processing of a vast amount of information.

In the method employed in the present study, with reference to the list of the companies dealing with the strategic items, acquired and managed by the Korea Institute of

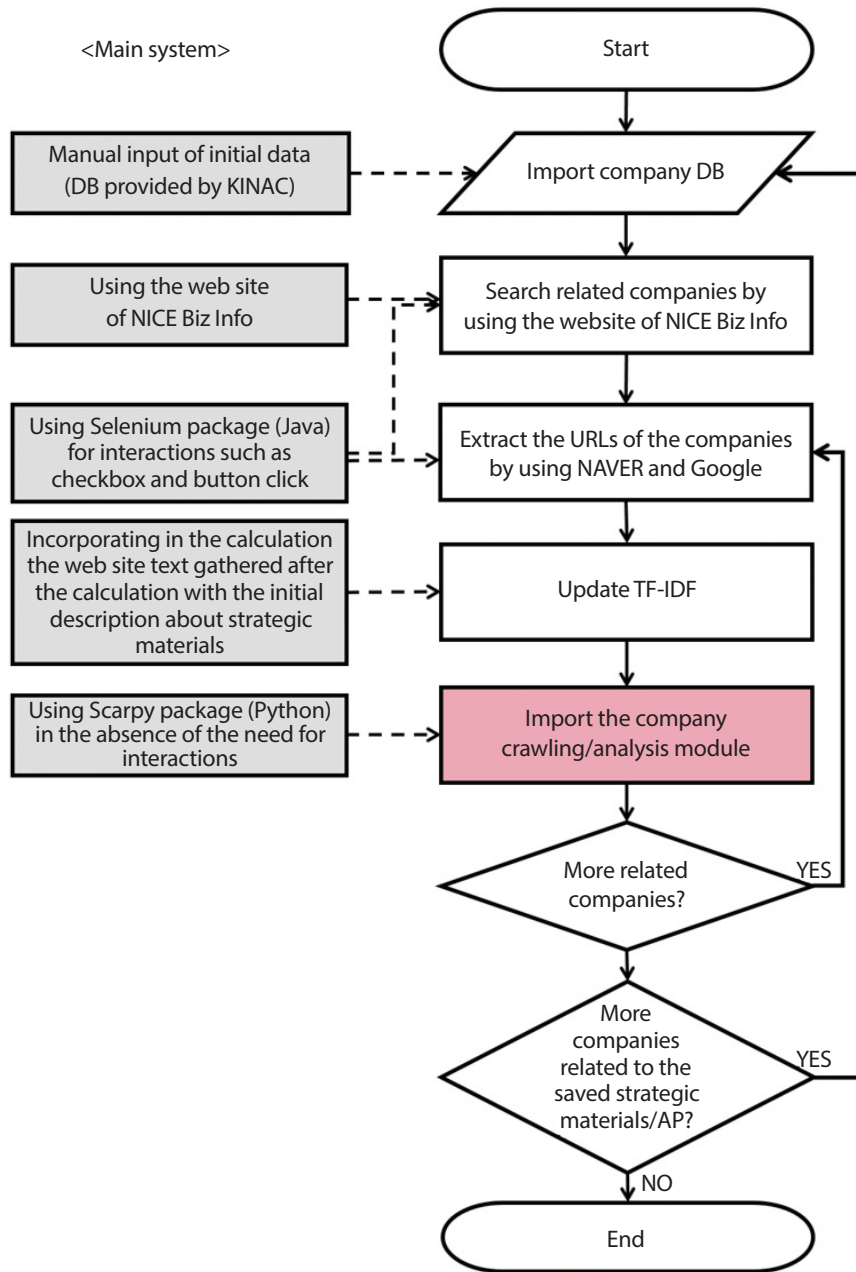


Fig. 1. The overall algorithm of the Nuclear Industry Information Gathering and Analysis System.

Nuclear Nonproliferation and Control, the companies that belong to a similar business category, have a similar scale, and deal with similar items in comparison with the companies that are already on the list are sequentially searched and analyzed. Several companies having a high relevance

with a company that deals with strategic items and that is already on the managed list are selected, and the official web sites of the selected companies are searched automatically. The information is then acquired from the official web sites through crawling. The information about similar

companies in South Korea may be obtained from the web site of the NICE Biz Info (<https://www.nicebizinfo.com/>). Other countries may have similar web sites that allow the users to query the analytical information about the business and industry and that provide the relevant data, and thus the algorithms and systems may be similarly applied. The system was configured to immediately transfer the results collected by the Gathering System to the analytical system for analysis. This approach shortens the time required for the users to view the initial information, compared to the method in which crawling is performed for all the companies subject to information gathering and then the entire information is transferred to the analytical system. Since the analytical results for each company are immediately provided, the user is given the time to search for further information (web site information in the form of flash or image files) that has not been acquired through the crawling and to determine whether the company deals with strategic items. Fig. 1 illustrates the overall algorithm of the Nuclear Industry Information Gathering and Analysis System.

### **3. Nuclear Industry Information Gathering System**

#### **3.1 Design of Information Gathering System**

The major purpose of the Nuclear Industry Information Gathering System is to periodically search for information about the nuclear power-related companies, access their official web sites, and acquire the industrial information. Acquiring the industrial information requires three external interfaces. As described above, the URLs to access the official web sites of the individual companies were obtained from the official web site of the NICE Biz Info (<https://www.nicebizinfo.com/>). Access to the official web sites of the individual companies allows for the extraction of text in all pages of the domain. Hence, the first external interface of the Information Gathering System is the official web site

of the NICE Biz Info. As will be described later, the official web site of the NICE Biz Info provides information based on the relatedness between companies, not the information about the official web sites of the related companies. Therefore, a separate task is needed to search for the official web sites of the companies acquired from the NICE Biz Info by using a search engine such as Google. Therefore, an additional external system interface with Google is included. Once the official web site of a company is found, one more external system interface is needed to directly approach the official web site of the company. However, the system interfaces, all of which are for the connection between the Nuclear Industry Information Gathering System and the web system, may be configured by a similar method. The gathered information is immediately transferred to the Nuclear Industry Information Analysis System, which will be described later. Therefore, an internal system interface should be considered for the connection with the Analysis System. Finally, since a user should be provided with the information on a screen, a user interface should be taken into consideration. Fig. 2 illustrates the structure and the internal and external interfaces of the Nuclear Industry Information Gathering System.

A data table was defined for the Nuclear Industry Information Gathering System. In the table, the meta-information about the collected companies is recorded. The names of the companies are saved in Korean and English with the official web sites, e-mail addresses, and telephone numbers. Since the Nuclear Industry Information Gathering System is applied only to Korean companies (under the Korean laws related to export and import), the result of whether the company is a Korean company is also saved.

When the system is operated, the system approaches the official web site of the NICE Biz Info to search for nuclear power-related companies saved in the database and acquires a list of the companies related to the individual companies in the database. As the system is continuously operated, the information about the searched companies is evaluated in terms of the relatedness with the strategic

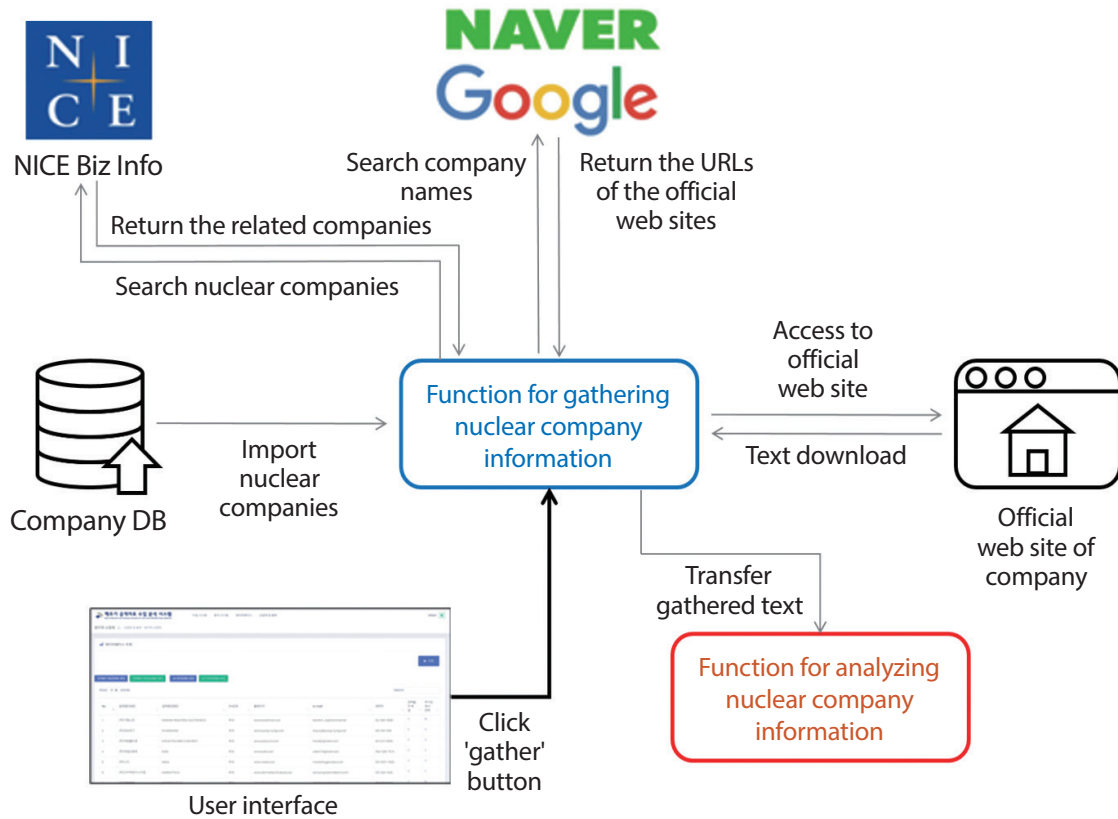


Fig. 2. Schematic of the structure and interfaces of the Nuclear Industry Information Gathering System.

items and the nuclear fuel period, and then registered to the company database. When the Information Gathering System is operated again the next time, the information is used again to search for the nuclear power-related companies. In summary, the Nuclear Industry Information Gathering System extracts the general information of the companies, such as the e-mail address and telephone number, from the official web sites of the companies, saves the extracted information in the database, and transfers the remaining text to the Analysis System.

### 3.2 Algorithm (Crawling)

As described above, the operation of the Information Gathering System is divided into the step of searching for the related companies and the step of gathering text from the official web sites of the individual companies. The first

step of searching for the related companies, based on the official web site of the NICE Biz Info, is carried out by a dynamic information gathering method that involves positive interactions with the components of the official web sites, including typing the company name into the search key word input window of the web site, clicking the search button, and selecting the check boxes and applying the search filters to obtain the search results. On the other hand, in the step of gathering the text, crawling is carried out by a static information gathering method, since crawling of the data can be performed simply by approaching the URL addresses and extracting the HTML source codes. Fig. 3 shows the crawling algorithm of the Nuclear Industry Information Gathering System.

A dynamic information gathering method is applied to the cases where the data can be viewed only after going through a separate procedure, such as login, or where the

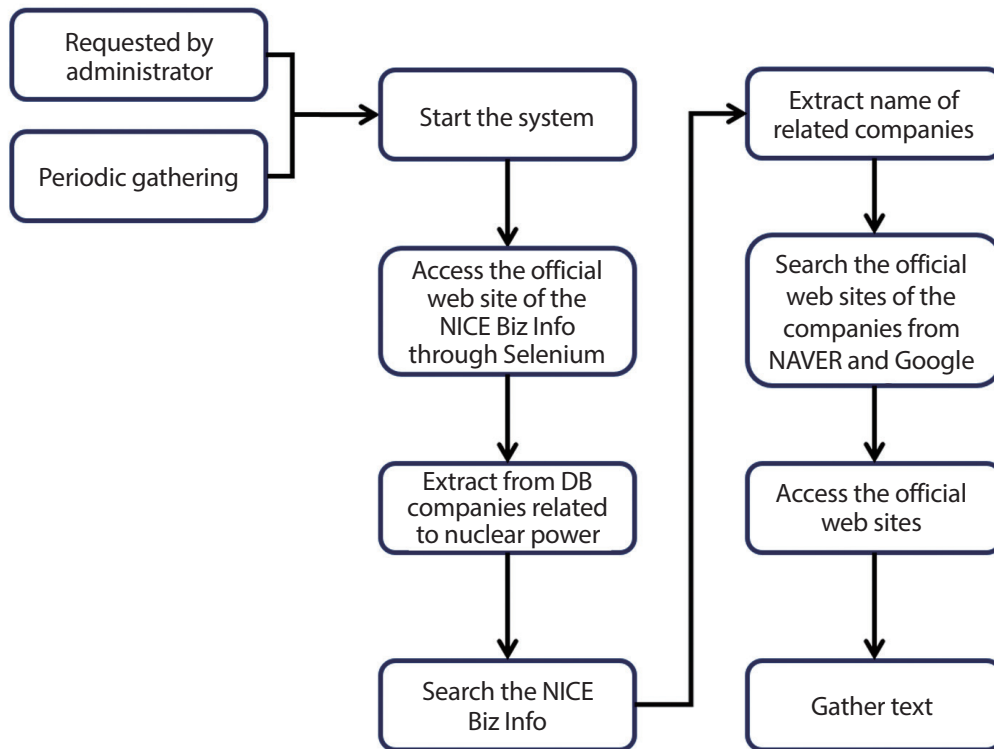


Fig. 3. Crawling algorithm in the Information Gathering System.

data inside the pages is changed on a real-time basis as in the Facebook and Instagram. A dynamic information gathering method employs the Selenium library [5], which can automate the web browser in the Java, Python, C#, Ruby, JavaScript, and Kotlin environments. The Selenium library can gather data by interpreting the sources codes even when the web components are dynamically rendered by JavaScript. However, the Selenium library requires a long waiting time for crawling until the HTML results are presented in the browser and all the web components become ready. A crawling test performed in the development environment based on the Selenium library showed that over one hour was required to process each official website of the companies. The Selenium library may be less effective in the step of gathering the text from the official web sites of the companies, but a dynamic information gathering method is necessary in the step of gathering the related companies to control the web components in communication with the

official web site of the NICE Biz Info. Hence, the system was configured by employing the Selenium library.

On the contrary, a static information gathering method is applied to the cases where the data can be viewed simply by entering the URL address without any other procedures or the data inside a page remain unchanged unless the page is refreshed. A static information gathering method employs the Scrapy library, which performs web crawling and scraping in the Python environment, requesting data and receiving the results simply through the URL address without a web browser [6]. Therefore, the Scrapy library is much faster than the Selenium library. A test performed in the Scrapy library development and operation environment showed that about one minute was required to process each official website of the companies, indicating that the Scrapy library is appropriate for the static information gathering method. Table 1 shows the difference between the Selenium and Scrapy libraries.

Table 1. Comparison of the characteristics of the crawling libraries

Library	Selenium	Scrapy (request)
Page access	Continuous access by using a browser	Single access through a URL address
Speed	Slow	Fast
Gathering capability	Virtually unlimited in the targets of information gathering	Applicable only to static pages
Gathering method	Dynamic gathering	Static gathering

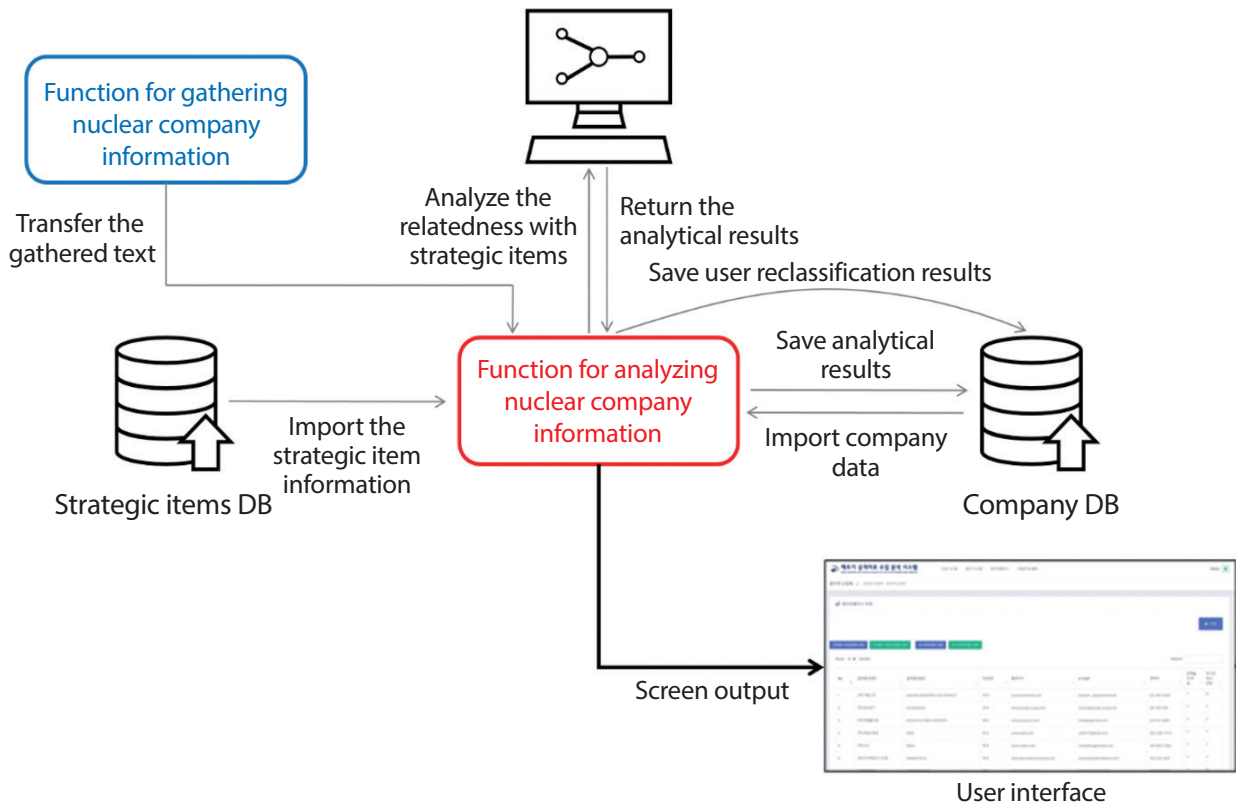


Fig. 4. Schematic of the structure and interfaces of the Nuclear Industry Information Analysis System.

## 4. Nuclear Industry Information Analysis System

### 4.1 Design of Analysis System

The Nuclear Industry Information Analysis System was designed to receive the gathered information from the Nuclear Industry Information Gathering System for automatic analysis. Fig. 4 is a schematic of the structure and interfaces

of the Nuclear Industry Information Analysis System. The Analysis System has only an internal interface with the Information Gathering System with no interfaces with other external systems. The user interface is also shared with the Information Gathering System. The DB table provides the fundamental data for determining whether the collected items belong to the strategic items, mostly the lists of the strategic items that the regulatory institutions have and the details stipulated by the laws.



Table 2. Detailed conditions for classifying strategic items and AP report targets (summary of five classification methodologies)

Classification conditions
1. Has a strategic item been found?
2. Is the found item not used in the general industry?
3. Has a strategic item been found?
4. Is there a name of a large category corresponding to the found item?
5. Has a strategic item been found?
6. Is there a name of controlled technology corresponding to the found item?
7. Has a strategic item been found?
8. Is there a name of classified technology corresponding to the found item?
9. Has a strategic item been found in the text?
10. Is the similarity between the description corresponding to the found item and the text above a certain level?

## 4.2 Realization of the Analysis System

Based on the text taken from the official web sites of the companies, the Analysis System determines whether the items dealt with by the companies are a strategic item or a nuclear fuel cycle-related item. Various determination conditions, based on the five classification methodologies (see Tables 2 and 3), are used to determine whether the gathered companies deal with a strategic item or a nuclear fuel cycle-related item and provide the determination results to the users. Even if the misclassification rate may be increased, to avoid missing any company that deals with a strategic item or a nuclear fuel cycle-related item, when the determination result acquired by at least one of the five classification methodologies corresponds to a strategic item or a nuclear fuel cycle-related item, the company is considered as one that deals with a strategic item or a nuclear fuel cycle-related item and recorded in the DB. To reduce the misclassification rate of the conservative determination system, the system was configured to allow an expert to verify the determination result and make the final decision regarding the companies that are initially considered by the Analytical System as dealing with a strategic item or a nuclear fuel cycle-related item.

The Analysis System reopens the text file saved in the Information Gathering System, reads each line of the sentences or paragraphs, and calculates the results by the five

methodologies.

The Analysis System compares the gathered text and the strategic item DB to determine whether a company deals with a strategic item or a nuclear fuel cycle-related item. The strategic item DB includes detailed information data about strategic items of all types, including the name, ID, strategic item code, and description. The IDs are 4-digit or 8-digit numbers, wherein the items of a 4-digit ID are the names of the large categories of the sub-items having an 8-digit ID. For example, light water nuclear reactor, heavy water nuclear reactor, reactor core, nuclear steam supply system, system-integrated modular advanced reactor, primary heat transport system, reactor cavity cooling system, etc. are included in the large category of “nuclear reactor”. Additionally, the strategic item DB may be used to verify whether each of the strategic items is applicable to the general industry.

The Analysis System searches the names of the strategic items in the text to classify the text. However, since the names of the strategic items generally consist of several words, even including stop-words such as postpositions, the names without modifications may not be effectively searched in the text. Hence, the system was designed to separate the individual words from the names of the strategic items and extract the nouns only for the search. In addition, because the words separated from the names of the strategic items may appear independently or not all the



Table 3. Criteria for classifying strategic items and AP report targets

Item search	Search of additional conditions
	Is the found item not used in the general industry?
Agitator(s)	Name of large category Uranium conversion factory and a system specially designed for that purpose [“uranium”, “conversion”, “factory”, “purpose”, “design”, “system”]
	Controlled technology Uranium conversion equipment [“uranium”, “conversion”, “equipment”]
	Classified technology Conversion [“conversion”]
Calculate the text cosine similarity of the description	

Table 4. Results of the system test

	Classification	Dealing with a strategic item	Dealing with no strategic item
Possible to analyze	Dealing with a strategic item (expected)	13	17
	Dealing with no strategic item (expected)	8	126
	Sub-total	21	143
	Accuracy	0.619	0.881
Impossible to analyze	Abnormal text extraction	7	51
	Official web site inaccessible	0	2
	Official web site absent	3	14
	Sub-total	10	67
	Total	31	210

words may appear in the text, the item search condition was set to search the words that appear over a certain level within a predetermined range. For example, the item name ‘reactor cavity cooling system’ is searched not as a whole but as four different words: ‘reactor’, ‘cavity’, ‘cooling’, and ‘system’.

All five text classification criteria begin with the assumption that the word has been found. The first condition is to verify if the found item is also used in the general industry. If not, the classification continues by checking the remaining four conditions. The second condition is whether there is a name of a large category corresponding to the found item. The third and fourth conditions are whether the two technology names are additionally included in the text. If none of the four conditions is satisfied, the Analysis System finally calculates the text cosine similarity of the

description of the item and the gathered text. If the cosine similarity is over a predetermined level, the company is classified as one that deals with a strategic item or a nuclear fuel cycle-related item. The cosine similarity is calculated by using the Term Frequency - Inverse Document Frequency (TF-IDF) value of each word. More specifically, the two texts are converted into one-dimensional vectors having the TF-IDF values of the words as elements, and the vectors are subject to an inner product to calculate a cosine function with respect to the angle between the two vectors. Since the TF-IDF value of a same word is dependent upon the text, only the IDF value is recorded in the DB. Each time when it is necessary, the TF-IDF value is calculated by counting the number of appearances of the word.

Table 4 shows the results of a test with the Nuclear Industry Information Gathering and Analysis System. The

analysis of the test results showed that the classification model rightly classified 13 of the 21 companies that actually deal with a strategic item, and also rightly classified 126 of the 142 companies that actually deal with no strategic item. In Table 4, the classification accuracy shows 62% and 88% for the companies dealing with and not dealing with a strategic item, respectively.

The eight companies that were classified as dealing with a strategic item though not doing so actually were mostly the cases where the valid information was not extracted from the official web sites for comprising information in the form of image or video. The companies classified into the category of 'Impossible to Analyze' corresponded to the cases where the text was not extracted normally through the information gathering module or the official web site was inaccessible. Among them, the seven companies corresponded to the cases where the official web site was based on Adobe Flash, which did not allow for the extraction of the text by means of the Scrapy model. In order to improve classification accuracy, it is necessary to improve the method of collecting text on the web site of the 'Impossible to Analyze' category above mentioned.

## 5. Denied Persons Gathering System

As mentioned earlier in this article, 'denied persons' refer to entities or individuals with whom the trade is limited or should be carried out with caution to seek international security and world peace, and the technical review for the export license requires to verify whether the export target is on the Denial Lists or not. Currently, the Korean Security Agency of Trade and Industry provides the Denial Lists, but it fails to reflect the frequently updated list. Hence, a system is necessary to gather the information from the official websites of the government and the relevant institutions that provide the Denial Lists regularly and update them [7]. When an item to be exported by an exporter is determined as a non-strategic item, to determine whether the item is

subject to the catch-all licensing, the exporter may use the online civil application system to check whether the buyer, final cargo receiver, and end-user are on the Denial Lists designated by the international community. The export license examiner may use the online civil application system in the review of the technology for licensing to check whether the export target is on the Denial Lists [8].

The Denial Lists are provided by the UN Security Council and the US, UK, and Japanese governments. The key function of the system of the present study is to access these web sites and regularly download the lists.

The Denial List gathering system regularly accesses the web sites that provide the Denial Lists, downloads the list files, and performs parsing of the character strings in the files. The Denial Lists from the UN and other countries are provided in various formats including PDF, CSV, XML, HTML, JSON, and Excel. Among the file formats provided by the individual web sites, the file formats for the present system were determined by considering the convenience of parsing and the quality of the parsing results. The gathered data are entered into the DB to replace the existing data, and presented on the screen according to the conditions or styles required by a user. Fig. 5 shows the overall procedures of the Denial Lists Gathering System described above. A dynamic gathering method based on the Selenium library is used to download the Denial Lists. The institutions that provide the Denial Lists and their web sites are shown in Table 5.

The Denial List Gathering System was configured to not only automatically update the data from the four Denial Lists but also enable the administrator to update the desired data. The basic screen provides the necessary information including the name, reference number, classification, and affiliated country so that the civil applicant and the regulator may utilize the information about the Denied Persons. The detailed information screen provides most of the information about the Denied Persons, including the address, place of birth, nickname, dates of starting and finishing the sanctions, and the description of the person or entity.

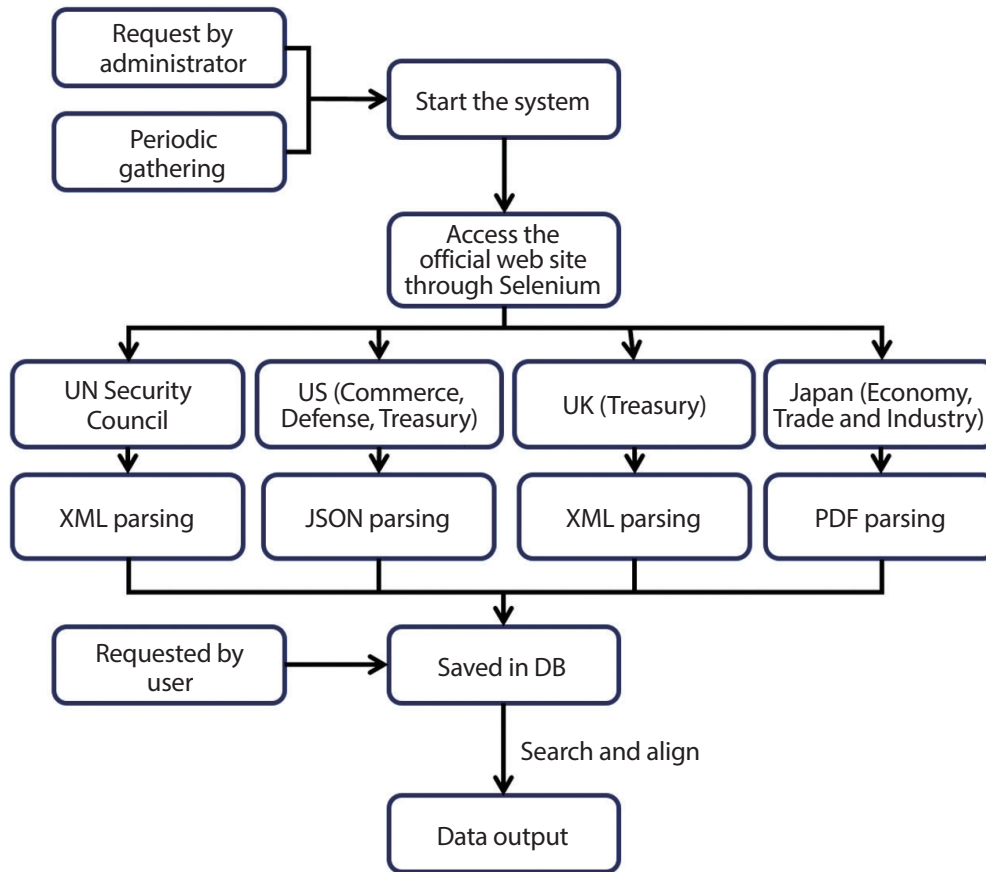


Fig. 5. Denied persons typified by the Denial List gathering system procedure parsing.

Table 5. Institutions providing Denial Lists and their web pages

Institution	Web page
UN Security Council	<a href="https://www.un.org/securitycouncil/content/un-sc-consolidated-list">https://www.un.org/securitycouncil/content/un-sc-consolidated-list</a> [9]
US Federal Commerce, State, and Treasury	<a href="https://www.trade.gov/consolidated-screening-list">https://www.trade.gov/consolidated-screening-list</a> [10]
UK Treasury	<a href="https://www.gov.uk/government/publications/financial-sanctions-consolidated-list-of-targets/consolidated-list-of-targets">https://www.gov.uk/government/publications/financial-sanctions-consolidated-list-of-targets/consolidated-list-of-targets</a> [11]
Japanese Ministry of Economy, Trade and Industry	<a href="https://www.meti.go.jp/english/policy/external_economy/trade_control/index.html">https://www.meti.go.jp/english/policy/external_economy/trade_control/index.html</a> [12]

## 6. Conclusions

Companies dealing with a Trigger List item needs to understand the export control system for the strategic items to avoid being involved in illegitimate export of the

items. This study was conducted to develop the Nuclear Industry Information Gathering and Analysis System to discover the companies having a low level of awareness about the export control system for the strategic items and incorporate them into the system. The Information

Gathering System was developed by establishing a database regarding the companies that deal with the Trigger List items, and the database was used as the initial input key words for the Information Gathering System. The Information Gathering System was configured to acquire the names of the companies similar to the initial input key words from the web site that provides the business information, access the official web sites of the companies through a search engine, and perform crawling to gather information. A database regarding the Trigger List items was established to comparatively analyze whether the gathered companies deal with a strategic item. The Analysis System determines whether the companies deal with a strategic item based on five classification methodologies. The classification accuracy was 62% for the companies dealing with a strategic item, and 88% for the companies dealing with no strategic item. Most of the errors in the classification of the companies dealing with a strategic item corresponded to the cases where the analysis was impossible as the information was difficult to extract from the official web sites of the companies. Therefore, the system needs to be improved to extract information from the web sites comprising images and other formats. The Denial List Gathering System was developed to periodically access the web sites of the UN, the US governments, and the like that provide the relevant information, gather the updated Denial Lists, and provide them to the examiners of the export license. Interlocked with the online civil application system for export control, the Denial List Gathering System is used in the examination of the export license to determine whether an export target is on the Denial Lists and to support the autonomous review by the exporters.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## Acknowledgements

This work was supported by the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS) using the financial granted by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea (No.1804021).

## REFERENCES

- [1] The National Assembly of the Republic of Korea, Foreign Trade Act, Act No. 16929 (2020).
- [2] The Ministry of Trade, Industry, and Energy, The Public Notice on Export Control of Strategic Items, Notice No. 2020-94 (2020).
- [3] P. Andrews-Speed, "South Korea's Nuclear Power Industry: Recovering From Scandal", *J. World Energy Law Bus.*, 13(1), 47-57 (2020).
- [4] S.J. Lee and S.J. Kim, "The Nuclear Industry Information Collection and Analysis System", *Proc. of the Korean Radioactive Waste Society Spring Meeting*, vol. 19(1), 55, June 2-4, 2021, Daejeon.
- [5] Selenium. "The Selenium Brower Automation Project." Selenium homepage. Accessed Jul. 20 2020. Available from: <https://www.selenium.dev/documentation>.
- [6] Scrapy. "Scrapy 2.4 Documentation." Scrapy homepage. Accessed Jul. 20 2020. Available from: <https://docs.scrapy.org/>.
- [7] Yestrade. "Denail List & Search." Yestrade homepage. Accessed Jan. 15 2021. Available from: <https://www.yestrade.go.kr/user/main.do?method=main>.
- [8] S.J. Lee and S.J. Kim, "Development of a Collection System to Provide Information of Denial List", *Proc of the Korean Nuclear Society Spring Meeting, KNS (P)-001-2021, P011B14, May 13-14, 2021, Changwon*.
- [9] United Nations. "United Nations Security Council Consolidated List." United Nations homepage. Accessed

- Feb. 8 2021. Available from: <https://www.un.org/securitycouncil/content/un-sc-consolidated-list>.
- [10] International Trade Administration. “Consolidated Screening List.” International Trade Administration homepage. Accessed Feb. 8 2021. Available from: <https://www.trade.gov/consolidated-screening-list>.
- [11] Government UK. “Guidance: Financial Sanctions Targets: List of All Asset Freeze Targets.” GOV.UK. homepage. Accessed Feb. 8 2021. Available from: <https://www.gov.uk/government/publications/financial-sanctions-consolidated-list-of-targets/consolidated-list-of-targets>.
- [12] Ministry of Economy, Trade and Industry. “Trade Control.” METI homepage. Accessed Feb. 8 2021. Available from: [https://www.meti.go.jp/english/policy/external\\_economy/trade\\_control/index.html](https://www.meti.go.jp/english/policy/external_economy/trade_control/index.html).