

## What Experts Care About in Satellite Observation : A Quantitative Analysis for Countering Nuclear Proliferation

Jae-Jun Han <sup>a\*</sup> and Nam Kyung Kim <sup>a</sup>

<sup>a</sup> Center for Nuclear Nonproliferation Strategy & Technology, Korea Institute of Nuclear Nonproliferation and Control,  
1418 Yuseong-daero, Yuseong-gu, Daejeon 34101, Republic of Korea

\*Corresponding author: jjhan@kinac.re.kr

### 1. Introduction

For countering nuclear proliferation, international society pays much attention to the potential or identified nuclear proliferant states. As providing advantages on accessibility to suspicious facilities in the rogue states, satellite observation has been playing an essential role in monitoring their potential nuclear activities.

Nuclear weapons as strategic arms have to consist of three components: fissile material, delivery system, and detonation devices. The weapons-grade fissile material can be acquired from uranium enrichment (about 90% U-235) or reprocessing used nuclear fuel (about 93% Pu-239). Hence, it is necessarily led to keep eyes on the relevant facilities and carriage of required materials by means of satellite surveillance. As for the delivery system, test sites and manufacturing factories are a matter of concern. Due to the size of a nuclear weapon, the area of interest for the delivery system is rather clearly to be revealed and investigated. On the other hand, there is no need to be special-exterior structures to develop detonation devices. Although any ambiguous activities at the border, e.g., trade traffic, could be captured on satellite imagery, a detailed investigation is under the implementation of strategic trade controls.

Identifying relevant activities not only for fissile material but also for delivery system involves expertise in the field of both nuclear fuel cycle and missile technology. Thus, the imagery analysts who are proficient in both areas, as well as remote sensing, are understandably small in number. They intermittently provide interpretation from satellite-observed information in the form of web-based articles via non-profit organisations such as 38 North and Center for Strategic and International Studies (CSIS). The articles inevitably include the logical basis for inferring nuclear proliferation, even though it is an external sign. As shown in Fig. 1, the annotations on satellite imagery represent what the experts practically capture, and the body contents of the article expand their interpretation.

This study quantitatively analyses the articles via web scraping to investigate what the experts care about in satellite observation. The following section presents the web scraping method for producing the article data set. Based on the data set, Section 3 discusses results from the article categorisation and the image annotation classification. The conclusions are drawn with future work in Section 4.



Fig. 1. Building construction continues in front of the ELWR (Excerpt from the 38 North article on 4 April 2018 [2])

### 2. Scope and Method

The web scraping is the process of automated gathering of data from the web pages which arrange objects in a more human-readable format [1]. Visiting every page and storing the hundreds of text contents and images are economically unfavourable. Thus, web scraping algorithms have been developed with the Python programming language [4] for the websites of interest: (1) 38 North [2] and (2) CSIS Beyond Parallel [3] which are the most active and being cited by the media. Both provide not only imagery-based articles but also text-only articles for various topics, and this study focused on the former only.

In terms of the entire platform of Python programming, the Selenium with Chrome web driver [5] is utilised to navigate webpages, and a parse tree for the pages that can be used to extract data from HTML is created via the BeautifulSoup library [6].

The scraping targets of an article are an article title, author(s), release date, body text, image(s), and image title(s). When scraping data from a website, targets are not always embedded in the identical HTML tag, so-called anomaly, which is the chronic challenge in the web scraping. For instance, the HTML tag content for the author(s) name occasionally is occupied as the organisation name instead, and the author(s) name is presented in the first line of the body text. Depending on the consistency of the first line, the automated gathering of the author(s)'s name may or may not be possible. If the number of the case is very few, filling

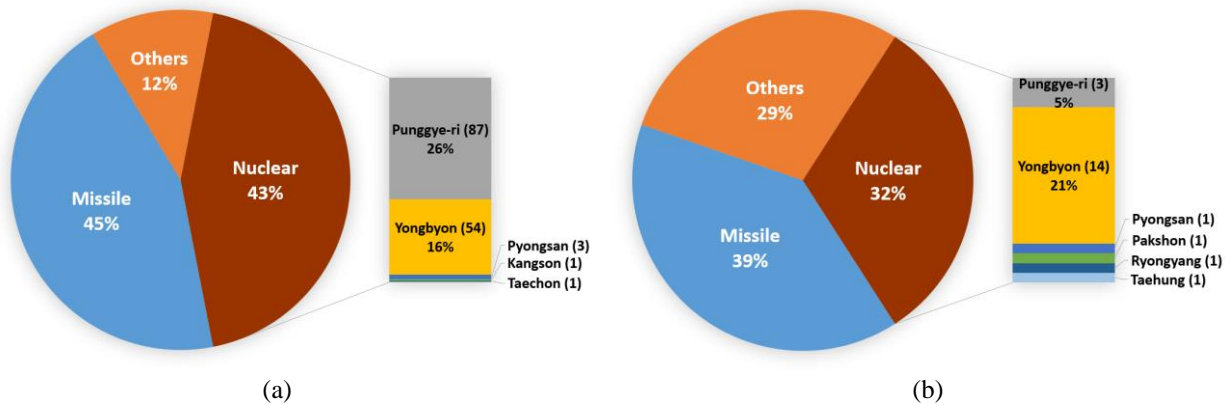


Fig. 2. Pie charts showing categorisation results for (a) 38 North (from November 2011 to June 2020) and (b) CSIS Beyond Parallel (from April 2016 to June 2020).

Table 1. The article information collected by the web scraping for the 38 North website from November 2011 to June 2020.

No.	Region	Category	Date	Title	Authors	Content (Article Text)	Img_Titl	No. of Images	Image1	Image39
1	Pyongyang	other	2020-06-09	Construction in Pyongyang: The G	['Martyn Williams', 'Peter M	New commercial satellite imag	['Figure 1. Features	8	24x453.jpg	
2	Pyongyang	other	2020-06-08	Progress Made on Mirim Parade T	['Martyn Williams', 'Peter M	New commercial satellite imag	['Figure 1. Overview	9	24x768.jpg	
3	Yalu River	other (China)	2020-05-29	New Yalu Bridge Construction Hits	['Martyn Williams', 'Peter M	Asphalt laying is complete on t	['Figure 1. Sinuiju-b	6	24x768.jpg	
4	Sinpo	missile	2020-05-29	Unusual Object at the Sinpo Secur	['Jack Liu', 'Peter Makowsky']	Recent reports indicate that Sc	['Figure 1. Overview	1	24x986.jpg	
5	Wonsan-Kalma	missile	2020-05-19	Wonsan-Kalma Beach Resort: Red	['Peter Makowsky']	Despite North Korea's previous	['Figure 1. Overview	10	24x768.jpg	
6	Pyongyang	other	2020-05-13	New Construction at the Mirim Pa	['Martyn Williams', 'Peter M	Commercial satellite imagery f	['Figure 1. Overview	3	24x768.jpg	
7	Pyongyang	other	2020-04-29	Construction of Pyongyang Genera	['Martyn Williams', 'Peter M	Commercial satellite imagery f	['Figure 1. Overview	5	24x768.jpg	
8	Wonsan-Kalma	missile	2020-04-29	Imagery Shows Train at Kim Jong U	['Martyn Williams', 'Peter M	Commercial satellite imagery f	['Figure 1. Train pre	3	24x768.jpg	
9	Wonsan-Kalma	missile	2020-04-25	Train Spotted at Kim Jong Un's Wc	['Martyn Williams', 'Peter M	A train probably belonging to K	['Figure 1. Overview	5	24x768.jpg	
10	Wonsan-Kalma	missile	2020-04-24	The Dual-Use Wonsan-Kalma Airfir	['Peter Makowsky']	Recent commercial satellite im	['Figure 1. Aircraft s	9	58x576.jpg	
11	Tanchon	other (hydroelectric)	2020-04-24	Tanchon Power Station Project: A	['Peter Makowsky']	Construction on the highly-tou	['Figure 1. Little cha	7	19-DG.jpg	
12	Pyongyang	other	2020-04-14	Pyongyang General Hospital After	['Martyn Williams', 'Peter M	Construction on the new Pyong	['Figure 1. Compari	5	24x768.jpg	
13	Pyongyang	other	2020-04-10	The Pyongyang General Hospital S	['Martyn Williams', 'Peter M	Commercial satellite imagery f	['Figure 1. Overview	3	24x768.jpg	
14	Punggye-ri	nuclear	2020-04-10	North Korea's Punggye-ri Nuclear	['Frank Pabian', 'Jack Liu', 'Pe	Commercial satellite imagery c	['Figure 1. Personne	13	58x577.jpg	
15	Sinpo	missile	2020-04-08	North Korea's Sinpo South Shipyar	['Peter Makowsky', 'Jack Liu']	Commercial satellite imagery c	['Figure 1. Indicato	4	24x922.jpg	
16	Pyongyang	other	2020-04-03	Construction Progressing Rapidly	['Martyn Williams']	Construction of the new Pyong	['Figure 1. Progress	7	57842.jpg	
17	Wonsan-Kalma	missile	2020-04-02	Wonsan-Kalma Beach Resort: April	['Peter Makowsky']	Recent commercial satellite im	['Figure 1', 'Figure 2	8	24x768.jpg	
18	Sohae	missile	2020-03-31	Sohae Satellite Launching Station:	['38 North']	A 38 North exclusive with anal	['Figure 1. Overview	11	24x768.jpg	
19	Pyongyang	other	2020-03-26	Construction Begins on New Pyong	['38 North']	A 38 North exclusive with anal	['Figure 1. Construc	3	24x703.jpg	
20	Yongbyon	nuclear	2020-03-20	More Underground Facilities Near	['Jacob Bogle']	Commercial satellite imagery r	['Figure 1. Overvie	7	24x694.jpg	
21	Yongbyon	nuclear	2020-03-13	North Korea's Yongbyon Nuclear C	Peter Makowsky, Olli Heino	A 38 North exclusive with anal	['Figure 1. Overview	7	24x768.jpg	
22	Sinpo	missile	2020-02-20	North Korea's Sinpo South Shipyar	['38 North']	A 38 North exclusive with anal	['Figure 1. Overview	3	24x769.jpg	
23	Yongbyon	nuclear	2020-02-14	North Korea's Yongbyon Nuclear C	Peter Makowsky, Frank V. Pe	A 38 North exclusive with anal	['Figure 2', 'Figure 1	8	24x768.jpg	
322	Sohae	missile	2012-03-28	Getting North Korea's Rocket to th	['David Wright']	I have been looking at options	[]	4	0x260.jpg	
323	Yalu River	other (China)	2012-02-19	China's Embrace of North Korea: T	['38 North']	Background In early June 2011	[]	11	2_Pic1.jpg	
324	Yongbyon	nuclear	2011-11-14	North Korea Makes Significant Pro	Robert Kelley and Mehdi Sar	This article is a "38 North" exc	Figure 1. One of the	7	11_fig2.jpg	

Table 2. The article information collected by the web scraping for the CSIS Beyond Parallel website from April 2016 to June 2020.

No.	Region	Category	Date	Title	Authors	Content (Article Text)	Img_Titl	No. of Images	Image1	Image47
1	Hungnam	other	2020-06-16	Kim Zeroes in on Fertilizer Produc	['Joseph Bermudez', 'Victor	Key Findings At Kim Jong-un's	['An October 30, 20	3	5-09-01.jpg	
2	Pyongsan	nuclear	2020-05-29	Pyongsan Uranium Concentrate P	['Joseph Bermudez', 'Victor	Overview of the southern sect	['Overview of the s	4	kups_2.jpg	
3	Sil-li	missile	2020-05-05	Sil-li Ballistic Missile Support Facil	['Joseph Bermudez']	Key Findings A new facility is r	['Overview image c	19	ility-01.jpg	
4	Manpo	other(China)	2020-05-04	Making Solid Tracks: North Korea	['Joseph Bermudez', 'Marie	New Original Analysis South K	[]	7	581630.jpg	
5	Yongbyon	nuclear	2020-03-11	Yongbyon Declassified Part VII: Fi	['Joseph Bermudez']	KH-4B Mission 1116, April 20,	['An overview of th	6	_VII-01.jpg	
6	Yongbyon	nuclear	2020-02-14	Yongbyon Update: February Mov	['Joseph Bermudez', 'Victor	Key Findings Satellite imagery	['Three specialized	1	2-14-01.jpg	
7	Yongbyon	nuclear	2020-02-11	Yongbyon Update: February Mov	['Joseph Bermudez', 'Victor	Key Findings Satellite imagery	['Three specialized	1	2-10-01.jpg	
8	Ippul-tong	missile	2020-01-14	Unique Facility in the Sanum-don	['Joseph Bermudez', 'Victor	Key Findings A new facility in	['Overview of the I	14	ility-06.jpg	
9	Chamjin-ni	missile	2019-12-19	December 2019 Update: Chamjin	['Joseph Bermudez', 'Victor	Key Findings The Chamjin-ni v	['Overview of the C	3	2-12-01.jpg	
10	Sinpo	missile	2019-12-17	December 2019 Sinpo Update No	['Joseph Bermudez', 'Victor	Key Findings As with Beyond	['Overview of the S	1	9-01-1.jpg	
11	Tonghae	missile	2019-12-17	December 2019 Update: Tonghae	['Joseph Bermudez', 'Victor	Key Findings The Tonghae Sat	['Overview of the T	8	2-16-01.jpg	
12	Nampo	missile	2019-12-14	December 2019 Update: The Nam	['Joseph Bermudez', 'Victor	Key Findings The submersible	['Minor activity is c	1	2-02-01.jpg	
13	Magunpo	missile	2019-12-13	December 2019 Update: The Mag	['Joseph Bermudez', 'Victor	Key Findings The Magunpo So	['Overview of the N	4	2-06-01.jpg	
14	Sohae	missile	2019-12-11	Sohae Engine Test Part of Coerciv	['Joseph Bermudez', 'Victor	Key Findings North Korea test	['Overview of the S	5	2-10-01.jpg	
15	Sinpo	missile	2019-12-04	December 2019 Sinpo Update: Nd	['Joseph Bermudez', 'Victor	Key Findings There is no evid	['Overview of the S	6	2-03-01.jpg	
16	Taehung	nuclear	2019-11-19	Mining North Korea: Magnesite P	['Joseph Bermudez', 'Marie	The Taehung Youth Hero Min	['The Taehung You	38	8_Map.jpg	
17	Yongbyon	nuclear	2019-11-14	Yongbyon Update: November Md	['Joseph Bermudez', 'Victor	Key Findings Recent satellite i	['Overview of the Y	7	1-09_1.jpg	
18	Yongbyon	nuclear	2019-10-22	Yongbyon Declassified Part VI: Im	['Joseph Bermudez']	KH-4B Mission 1109, March 17	['An overview of th	6	new-01.jpg	
19	Punggye-ri	nuclear	2019-10-17	Punggye-ri Nuclear Test Site: Im	['Joseph Bermudez', 'Victor	Analysis Analysis of satellite in	['Overview of the P	8	date_1.jpg	
20	Sinpo	missile	2019-10-09	Sinpo Update: Significant Conceal	['Joseph Bermudez', 'Victor	Key Findings The October 2, 2	['Overview of the S	8	oard-6.jpg	
21	Yongbyon	nuclear	2019-09-25	Yongbyon Declassified Part V: Exp	['Joseph Bermudez']	KH-4B Mission 1105, Novemb	['Overview of north	6	w-01-1.jpg	
22	Kumchon	missile	2019-09-06	Undeclared North Korea: The Kum	['Joseph Bermudez', 'Victor	Key Findings Located 1,100 kil	['rview of the Kum	23	1-ni_01.jpg	
23	Sinpo	missile	2019-08-28	Sinpo South Shipyard: Constructi	['Joseph Bermudez', 'Victor	Key Findings New Beyond Par	['Overview of the S	7	s_01-1.jpg	
63	h River, Hoeryong,	other	2016-11-02	The Politics of the North Korean F	['Andrew Natsios']	Between August 29 and Septe	[]	0		
64	h River, Hoeryong,	other	2016-11-02	Implications of North Korean Flood	[]	DPRK Flood Damage May Prec	['— Choi Song Min,	12	016-01.png	
65	u, Yalu River, (Dan	other(China)	2016-04-22	Decrease in Trade After North K	[]	Imagery Suggests Downtick in	[]	13	5_05_1.jpg	

up can be a wise option. The number of anomalies in this study is acceptable to manual handling. The release dates are converted into the international format (yyyy-mm-dd) for a managerial purpose.

The satellite imagery within the article is the ground for interpretation of potential nuclear activities, and more importantly, the annotation on the imagery directly indicates key messages of satellite observation. In this regard, during the web scraping, the imagery is renamed, including the release date (of the article), numbering in ascending order, and original file name. The stored image files are linked to the data cell in the MS Excel so that a user can easily access to open each image. Additionally, each caption (title) of the imagery is also collected if available. The data set is stored in the Python serialisation format (\*.pk) as well as the Microsoft Excel documents format (\*.xls).

The data collected with the refinement facilitates the post-analysis: (i) categorisation of the articles, (ii) classification of the image annotations, (iii) author network analysis, (iv) natural language process of the body text, and (v) correlation analysis with other sources. This study addresses the former two topics in the subsequent section.

### 3. Results

#### 3.1 Statistics with Categorisation

Tables 1 and 2 show the results of web scraping for the 38 North (324 articles from November 2011 to June 2020) and CSIS Beyond Parallel (65 articles from April 2016 to June 2020), respectively. The relationship between region and activity associated with nuclear weaponisation has been historically identified. Thus, given the region information from the article title (with body text and imagery as necessary), each article has been categorised into three topics: nuclear, missile, and others. The others cover hydroelectric, coal, chemical, agriculture, China issues, etc.

Figure 2 illustrates the proportions of categorical data from the articles. For an article dealing with multiple topics, it has been counted as each topic. Both websites have published nuclear and missile topics at similar rates. The articles of the nuclear category occupy 43% and 32% for the 38 North and CSIS-BP, respectively. The Yongbyon and Punggye-ri articles predominate in the nuclear category. It is worth noting that each region has been suspected for a specific process in nuclear proliferation, and the minority such as Pyongsan (mining and milling) and Kangson (enrichment) is not of low importance. Owing to the fact that the Yongbyon nuclear complex has been well-known as having various nuclear fuel cycle facilities, it was chosen for analysing annotated imagery.

As for the type of satellite sensing, the electro-optical (EO), infrared (IR), and synthetic-aperture radar (SAR) sensors are technically available, and the articles for the Yongbyon nuclear complex consist of four IR images and all EO images for the rest.

Table 3. Classification of the annotations extracted from satellite imagery interpretation by the 38 North and CSIS Beyond Parallel.

Type of Sign	Class	Sub-Class
direct /indirect	construction	construction (incl. maintenance)
		excavation
		dismantled
		trench
	C-crane	
direct	pipeline	pipe
indirect	powerline	power line
indirect	transport	rail car/flat car
		truck
		(other) vehicle
indirect	dredge	dredge(r)
		D-crane
indirect	object	equipment (or supplies)
		material
		container/tank
		(other) object
indirect	fuel	coal
direct	operation	snowmelt
		foam/liquid
		steam
		stain
		flow
		thermal pattern (IR imagery)
indirect	etc.	personnel
		tent
		grain
		vegetation
		roof opened
	side section (of structure)	

#### 3.2 Classification of Activities from Satellite Observation

A total of 510 images from the web scraping has been analysed for annotation extraction. It aims at classifying what the experts practically capture so that the key indicators on the satellite observation can be identified.

Firstly, the annotation has been recognised with assistance from the Clova Optical Character Recognition (OCR) platform [7]. Even though the root meaning of the annotations is identical, it shows multifarious variations depending on the author(s) as well as the diversity of language expression. Thus secondly, the extracted phrases have been refined into simple forms; this process is the drafting of shared characteristics, resulting in the sub-class. Lastly, the (upper) class is created by finding common features of sub-class activities. Table 3 summarises nine classes and 29 sub-classes derived from the annotations of 510 images.

The class from the satellite observation indicates direct or indirect activities for potential nuclear

proliferation. The type of sign (direct or indirect) in the left column of Table 3 is decided for the classified activities based on connectivity to the facility for nuclear fuel cycle or weaponisation. For instance, class ‘operation’ stands for observation associated with fuel cycle facilities such as steam arise from the 5 MW<sub>e</sub> reactor, snowmelt on the uranium enrichment plant, etc. On the other hand, class ‘transport’ including rail car/flat car, truck, etc., represents shipment or (at least) movement within the nuclear complex, and they are not directly connected to the facility for nuclear fuel cycle or weaponisation.

#### **4. Conclusions**

Due to the deficiency of the subject-matter experts in both the imagery interpretation and nuclear fuel cycle, this study has built the data set from satellite-observed information via the web scraping of the articles for countering nuclear proliferation. Through the 38 North and CSIS Beyond Parallel websites, the expert(s) has produced 389 articles of which topics have been categorised into the nuclear, missile, and others. In the nuclear category, specifically, the proportions of Yongbyon and Punggye-ri are the highest. The data set is expected to be a cornerstone for the in-depth analysis, including author network analysis, natural language process, etc.

Further, 510 annotated images for the Yongbyon nuclear complex have been extracted to classify what expert(s) care about in satellite observation. The indicators for inferring nuclear proliferation have been classified into nine classes: construction, pipeline, powerline, transport, dredge, object, fuel, operation, etc.

The results for classification are valuable in that they can be utilised by non-experts or novice interpreters in countering nuclear proliferation. Moreover, for automated change detection with algorithms, this provides a framework for satellite-observed objects that the system has to detect.

#### **ACKNOWLEDGEMENT**

This work was supported by the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS), granted financial resource from the Nuclear Safety and Security Commission (NSSC), Republic of Korea (No. 1905009).

#### **REFERENCES**

- [1] R. Mitchell, Web scraping with Python: Collecting more data from the modern web, O’ Reilly Media, Inc., 2018.
- [2] Articles – Topic: Satellite Imagery, Available at <https://www.38north.org/topics/satellite-analysis/> (Accessed: 26 Aug 2020)
- [3] Articles – Imagery, Available at <https://beyondparallel.csis.org/imagery/> (Accessed: 26 Aug 2020)

[4] G.V. Rossum and F.L. Drake, The python language reference manual. Network Theory Ltd., 2011.

[5] Selenium WebDriver documentation, <https://www.selenium.dev/documentation/en/> (Accessed: 26 Aug 2020)

[6] BeautifulSoup documentation, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (Accessed: 26 Aug 2020)

[7] Naver Clova Optical Character Recognition (OCR) platform, <https://clova.ai/ocr> (Accessed: 26 Aug 2020)