# 1. Introduction

## 1.1 Purpose of this Document

This document serves as a guideline of the annotation for the text describing High-Pressure Gas accident situations. The purpose is to extract information from the accident case database created by The High Pressure Gas Safety Institute of Japan.

## 1.2 Database

The target of this annotation includes the reports of accidents related to high-pressure gas that have occurred in the past at factories handling high-pressure gas, such as chemical plants. The database can be downloaded from this website(https://www.khk.or.jp/public_information/incident_investigation/hpg_incident/incident_db.html).

## 1.3 Annotation Tasks

We carry out annotations for the following three types of Natural Language Processing (NLP) tasks.

- NER（Named Entity Recognition）
  Extract words corresponding to specific named entities such as equipment and chemical substances.
- CE（Cause and Effect）
  Extract text corresponding to the causes and results of the accidents.
- IR（Information Retrieval）
  Search for past accidents situations related to the keywords or text concerning operations.

## 1.4 Annotation Tools

For Named Entity Recognition (NER) and Cause and Effect (CE) annotations, we use brat (https://brat.nlplab.org/), an open-source annotation tool. For Information Retrieval (IR), we use Excel.

## 1.5 Procedure for Annotation

In order to maintain the quality of the annotations performed by three annotators, each task will be proceeded after we can confirm that the level of agreement (IRR: Inter-Annotator Agreement) among the three is at a certain standard.

## 2. NER（Named Entity Recognition）

### 2.1 Overview of NER

Extract potentially high-pressure gas accident-related words, such as chemical substances, from the target text.

### 2.2 Details of the NER Annotation

Extract words of the types (Entities) shown in Table 1.

Table 1． Entity with Descriptions and Examples

| | Entity | Descriptions | Examples |
|---|---|---|---|
| 1 | Products | Various gases.<br>Gaseous state at normal temperature and pressure.<br>Nouns.<br><br>※Do not tag items that are not general (things that do not appear even if you search the Web). | Mixed gas<br>Flammable gas<br>Refrigerant gas<br>Inert gas<br>Liquefied petroleum gas<br>Carbon dioxide gas<br>Sulphur dioxide gas<br>Liquefied petroleum gas<br>Freon<br>Hydrogen, Carbon monoxide, Acetylene, Methane, Ethylene |
| 2 | Chemicals | Chemical substances, reactants, and materials (other than gases) used in gas generation and process management。<br>Items not included in the above Products.<br>Nouns. | Water, water droplets, rainwater, wash water, hot water, pure water<br>H2O<br>Benzene<br>Austenitic stainless steel<br>Lubricating oil<br>C4-C6<br>Hydrocarbons |
| 3 | Storages | General equipment where above Products and Chemicals come into contact.<br><br>※Include equipment such as supports and insulators.<br>※Include expressions that indicate the entire plant or facility.<br>※ Do not include expressions indicating parts such as entrances and exits if they are placed at the end of a word. | Tank<br>Maturation furnace<br>Refining tower<br>Dehumidification tower<br>Separation tower<br>Heat exchanger<br>Piping<br>Valve<br>Gasket<br>Flange<br>BTX manufacturing equipment<br>Butadiene plant |
| 4 | Incidents | Incidents that resulted in or caused an accident, regardless of severity. | 破裂<br>爆発 |

| | | | |
|---|---|---|---|
| | | Include only incidents that actually occurred, and do not include situations that did not lead to an incident. | Seepage<br>Leakage<br>Fire<br>Serious injury<br>Death<br>Degradation<br>Concentration<br>Issuing of an alert, detection, awareness, (alarm) activation |
| 5 | Process | Handling of gas, and unit operations related to gas.<br>Abnormal processes are included in Incidents. | Filling<br>Distillation<br>Extraction<br>Reaction<br>Recovery<br>Mixing<br>Sealing<br>Nitrogen purge |
| 6 | Tests | Inspection devices and inspection actions outside the production process line.<br>Do not include inspection items such as XX concentration. | Inspection, visual inspection, three-month inspection<br>Detailed inspection, leakage inspection<br>Freon checker<br>Leak test<br>Analysis<br>Patrol |

2.3 Rules for Conducting Annotations

Tag the Cause and Effect in the text using the designation tool.

The following rules must be observed when tagging:

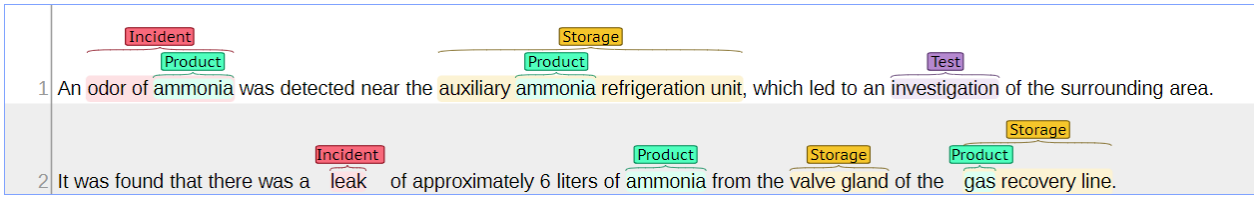
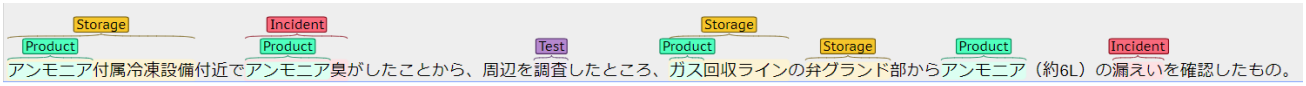
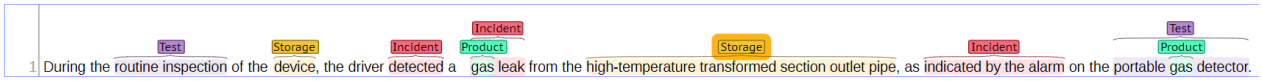
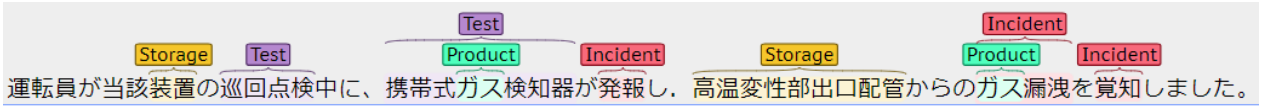
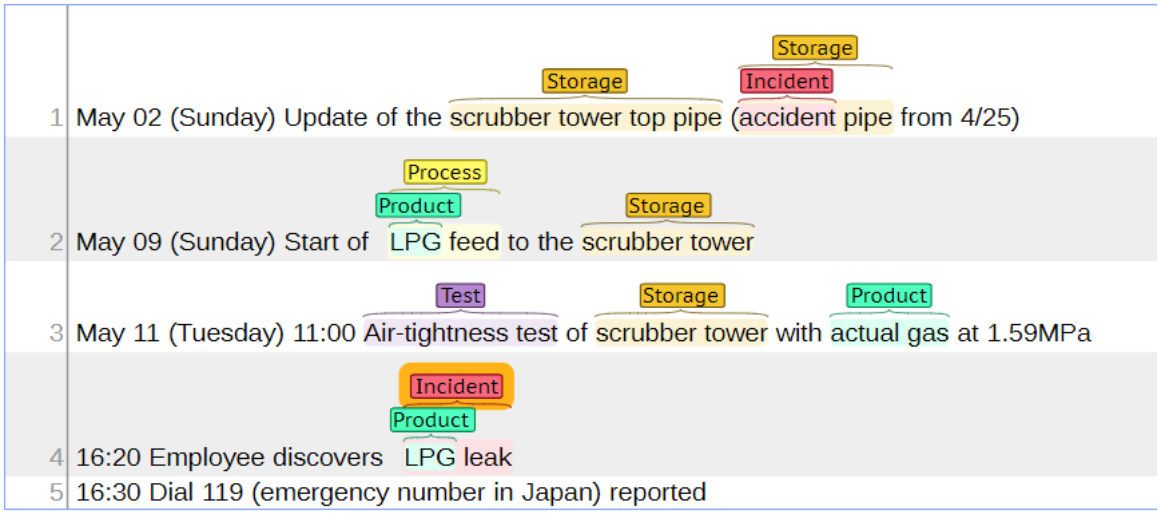
- Do not tag words indicating parts such as entrances, exits, and connections if they are at the end of a word. For example, tag "inlet pipe" as "inlet pipe", but for "heat exchanger outlet", only tag "heat exchanger".
- For tagging corresponding parts that indicate a range like "4-6", tag the entire "4-6".
- If the same word is used in different meanings, tag only the relevant entity.
- For words like "XX gas generation equipment," tag both Storage and Products (nested). For example, tag "XX gas generation equipment" as Storage and tag “XX gas”” as Products.
- If there is a modifier Process within Products, Chemicals, or Storage, do not tag Process. For example, do not tag "recycle" as Process in "recycle gas."
- Do not tag phrases containing particles like "of" in "XX of YY" (Tag only "XX" or "YY" separately).
- Tag abbreviations as well. However, do not tag specific abbreviations such as equipment or model numbers.
- Do not tag the state of individuals. Example: "Lack of perspective"

・Do not tag words within legal names or standards, such as the names of laws or regulations.
Example: "High-Pressure Gas Safety Act," do not tag "High-Pressure Gas."
・Only tag Pc when it pertains to gas handling.
Example: Do not tag "tightening further"

2.4 Examples of NER tagging

Examples shown in Table 2.

Table2. Annotation Examples

| | Examples |
|---|---|
| 1 |  <br>  |
| 2 |  <br>  |
| 3 |  |

| | |
|---|---|
| | 5月02日（日）洗浄塔塔頂配管の更新(4/25事故配管)<br>5月09日（日）洗浄塔へのLPGフィード開始<br>5月11日（火）11:00 洗浄塔実ガスによるI.59MPaでの気密テスト<br>16:20従業員がLPG漏えい発見<br>16:30 119番通報 |
| 4 | 1 During the operator's routine patrol, a leak was detected by the senses, and as a result of the investigation,<br>2 a leak was discovered from the flange of the heat exchanger.<br>3 By diffusing the combustible gas with steam blow, and while intensifying inspections, the leak was stopped by tightening.<br>4 The cause is presumed to be poor tightening management.<br><br>オペレーターの定期巡回中に五感により漏えいを感知し、調査の結果、当該熱交換器のフランジより漏えいを発見した。<br>スチーム吹付により可燃性ガスの拡散を行い、点検強化を図る中で増し締めにより、漏えいを停止させた。原因は、締付け管理不良と推定される。 |
| 5 | 1 On June 26th in the first year of Reiwa (2019), around 5:00 a.m.,<br>2 process gas was detected by the area monitor of the fluoropolymer manufacturing facility.<br>3 Upon inspection of the site, a leak of hexafluoropropylene was confirmed.<br><br>令和元年6月26日5時00分頃、フッ素樹脂製造施設のエリアモニターにてプロセスガスを検知し現場確認したところ、六弗化プロピレンの漏えいを確認した。 |

3. CE (Causes-and-Effects)

3.1 Overview of CE

Extract a sentence that represents the causal events (Causes) and its outcomes (Effects) that caused the accident.

3.2 Details of the CE Annotation

3.2.1 Overview of Effects

The accident outcomes (Effects) are classified as follows. It is assumed that ② and ③ occur only after the occurrence of ①. Please note that there may be samples that do not fall into any of these categories.

① Accidents primarily involving gas as the subject: The following two classifications:
- Event_Leak：Leakage of gas (here, gas = Products in NER)
- Event_others：Not related to leakage. Example: explosion, fire

② Damage_Property：Physical damage caused by accidents related to gas incidents

③ Damage_Human：Injuries caused by accidents related to gas incidents or physical damage

3.2.2 Overview of Causes

Extract the causal (cause-and-effect) sentences of the accident result (Effects). In samples where there is no corresponding sentence for Effects, there is also no Causes. Even if there is an Effect, there are also samples without a Cause statement.

3.2.3 Entity

Table 3 shows descriptions of each Entity and representative examples.

Table3. Entity and Descriptions

| | Entity | Descriptions | Examples |
|---|---|---|---|
| 1 | Event_Leak | Tag sentences in which gas leakage can be directly confirmed. However, automatic detection by equipment is not included due to the possibility of malfunction. Human detection is included. The definition of gas follows the NER Product. | Hydrogen and aniline leakage<br><br>(Not applicable example of sentence)<br>The leaked gas is hydrogen. |
| 2 | Event_others | Tag sentences containing accident events other than gas leakage. For example, explosions, fires, etc. | It is estimated that hydrogen, which has a low ignition energy, was ignited by static electricity. |
| 3 | Damage_Property | Tag sentences that confirm physical damage to equipment or facilities caused by Event_Leak and Event_others. Physical damage includes burst pipes, destruction of heat exchangers, etc. | Container ruptures. |
| 4 | Damage_Human | Tag sentences that confirm Huma n casualties caused by Event_Leak and Event_others and Damage_P roperty. Human casualties include deaths, injuries, and physical illnes | One employee injured left thigh and left ear. |

| | | | |
|---|---|---|---|
| | | ses. | |
| 5 | Cause | Tag sentences that confirm the event causing Event_Leak and Event_others. Target not only direct causes but also indirect causes (e.g., Cause's Cause)。<br>In case of ignition or explosion, the three elements of combustion (combustibles, oxygen, and heat) shall be noted cause. | As a result of reduced tightening torque in some of the flange sections cooled by hydrogen |

3.3 Rules for Annotation Implementation

Tag the Cause and Effect in the text using the designation tool.

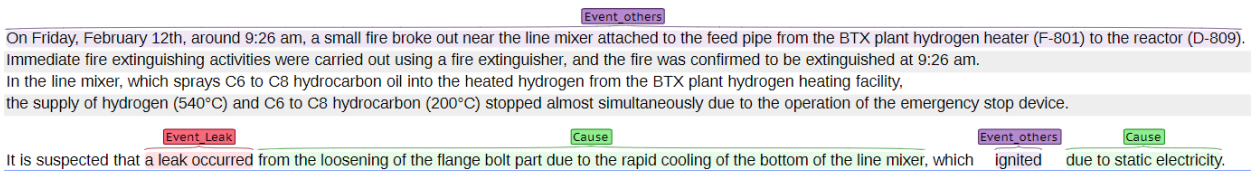
The following rules must be observed when tagging:

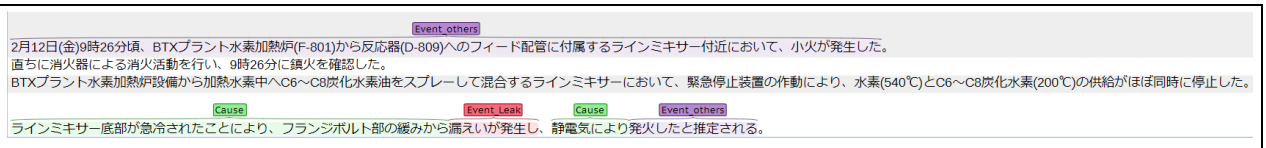
・Include in one sentence to be tagged:
  ・Who, When, Where, What
  ・Include endings up to verb phrases (e.g. Tag up to "leaked")
・Do not include in a sentence to be tagged:
  ・Punctuation at the end of tagging ", and."
  ・Such as "due to..."
  ・Conjunctions at the beginning of a sentence (e.g. And," "And then," "And then," etc.)
・How to separate each tagging
  ・Do not separate with "," but separate with ".".
  ・In the case of "broken and leaked," "broken" and "leaked" are two different tags, so separate them.
・No nesting.
・Do not tag the trigger for accident discovery (unless it is a causal factor in the accident)
  Example: "I noticed a strange odor,"

3.4 Examples of CE tagging

Examples shown in Table 4.

Table4. Examples of CE tagging

| No. | Examples |
|---|---|
| 1 |  |

| | |
|---|---|
| |  |
| 2 |  |

## 4. IR (Information Retrieval)

### 4.1 Overview of IR

Determine how highly related the text is between accident conditions for each accident.

### 4.2 Details of Annotation

#### 4.2.1 Labeling

Label the accident situation text for each accident according to the attributes in Table 5.

Table5. Attributes and Labels

| | Attributes | Labels | Descriptions |
|---|---|---|---|
| 1 | Types of high-pressure gas | a. Flammable (or flame retardant) gas<br>b. Toxic gas<br>c. Satisfies a and b<br>d. Not applicable | The high-pressure gas that caused the reported accident was classified from the perspective of danger in the event of an accident. Cases where the gas could not be identified were included under “d. Not applicable”.<br>The definition of flammable gas and toxic gas shall conform to the High Pressure Gas Safety Act in Japan. |
| 2 | Cause of accident | a. Equipment Factors | The events that caused or triggered the |

| | | | |
|---|---|---|---|
| | | b. Human Factors<br>c. External factors<br>d. Other factors | accident were classified. Equipment factors refer to those caused by initial defects in parts built into the equipment. Human factors refer to errors made in operation or judgment by people on site. External factors indicate those caused by events from outside the equipment, such as falling objects. |
| 3 | Accident Results | a. Leakage<br>b. Fires and explosions<br>c. a. and property damage<br>d. a. and human casualties<br>e. b and property damage<br>f. b and human casualties<br>g. Property damage and human casualties | The events that occurred as a result of the accident were classified. Physical and human damage were only considered if they occurred as secondary events, such as gas leaks or fires.<br><br>Property damage：Accidents resulting in damage to equipment or facilities due to fire or explosion<br>※Do not include damage to equipment or other items that caused the accident.<br>Human casualties：Accidents resulting in health hazards to humans due to leakage, fire, or explosion |
| 4 | Time span from cause to effect | a. Sudden<br>b. Long-term<br>c. Unknown | The classification was made based on the time from when the cause or trigger of the accident occurred until the accident event took place.<br><br>Sudden：Accidents where the results are caused generally within a few minutes to several tens of minutes from the occurrence of the cause. |
| 5 | Operational status of equipment at the time of cause occurrence | a. During steady-state operation<br>b. During non-steady state operation<br>c. During maintenance<br>d. Other situations. | The classification was made based on the operational status of the equipment at the time of the accident.<br>Non-steady state operation refers to operating conditions that differ from normal operation, such as immediately after the equipment starts running or during test operation002E |

4.2.2 Calculation of relevance

Based on the label information assigned to each text, calculate the relevance between texts. Adjust the weighting of each attribute according to the user performing the search task. Additionally, for certain attributes, there may be labels that are determined based on partial matches with the content, so caution should be exercised when adjusting the weighting in comparison to other attributes.

4.3 Examples of IR tagging

Examples shown in Table 6.

Table 6. Examples of IR (the alphanumeric in labels corresponds to Table 5)

| No. | Examples | labels |
|---|---|---|
| 1 | English:<br>On December 25, 2018, around 17:00, an abnormal alarm for the outdoor unit of the OPS plant air conditioning system was triggered, prompting us to shut down the equipment. An inspection by the manufacturer on January 10, 2019, revealed that the pressure in one of the two systems was zero, and 27.53 kg of refrigerant gas (R410A) had leaked from the evaporator in the indoor unit. It was deduced that the refrigerant side header (a vertically elongated shape) of the indoor unit evaporator (heat exchanger) had cracked due to repeated thermal contraction and expansion, leading to recurring stress at the tube plate contact point of the header connection pipe. It was confirmed that when the air volume is reduced for condensation prevention during operation, stress exceeding the standard is applied to the header connection pipe.<br><br>Japanese:<br>平成 30 年 12 月 25 日 17:00 分頃、OPS 工場空調機の室外機異常アラームが発生したため、機器を停止させた。平成 31 年 1 月 10 日にメーカーによる点検の結果、2 系統の内 1 系統の圧力が 0 であり、室内ユニット内の蒸発器からフロンガス(R410A）が 27.53 kg漏えいしたもの。室内ユニット蒸発器（熱交換器）のフロン側ヘッダ（縦長形状）が、熱収縮・膨張を繰返り返したっことにより、ヘッダ連絡管の管板接触部に繰り返し応力が発生し、割れが発生したものと推定した。結露防止のため風量を低下して運転する際、ヘッダ連絡管に基準を超える応力が働いていることが確認された。 | 1. d<br>2. a<br>3. a<br>4. b<br>5. a |
| 2 | English:<br>The flowmeter body joint part of the pipe section concerning consumption after pressure adjustment had minor leakage due to age-related degradation (23 years), which caused a decrease in surface pressure. The interlock was triggered by the most recent gas detector sensing (0.5ppm), which automatically stopped the supply and carried out hazard removal in the facility. It should be noted that there was no leakage outside the room. The cause is believed to be the decreased surface pressure of the flowmeter body joint due to age-related degradation (23 years). This was inferred from the results of the joint opening inspection of the part in question, which did not show any damage to the gasket part, but there was a thickness distribution (2-4mm).<br><br>Japanese: | 1. d<br>2. a<br>3. a<br>4. b<br>5. a |

| | | |
|---|---|---|
| | 調圧後の消費にかかる導管部の流量計本体継手部の経年劣化(23 年)により、面圧が低下し、微量の漏れが発生した。直近のガス検知機感知(0.5ppm)によりインターロックが作動し、自動的に供給閉止、並びに施設内の除害がなされた。なお、室外への漏えいはなかった。原因は、当該部の継手開放点検の結果、ガスケット部の外傷は認められず、厚みに分布(2～4mm)がみられたことから、経年劣化(23 年)により流量計本体継手の締結部の面圧が低下したためと推定される。 | |