Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021 - ISBN – 978-1-936263-34-9

Technical Language Processing for Efficient Classification of Failure Events for Safety Critical Equipment

Maria Vatshaug Ottermo[1], Solfrid Håbrekke[1], Stein Hauge[1], and Lars Bodsberg[1]

[1]*SESS, SINTEF Digital, Trondheim,7034, Norway*
maria.v.ottermo@sintef.no
solfrid.habrekke@sintef.no
stein.hauge@sintef.no
lars.bodsberg@sintef.no

ABSTRACT

The objective of this paper is to investigate how technical language processing can be utilized to automate classification of failure notifications for safety critical equipment on petroleum facilities. For such equipment, maintenance decisions are heavily influenced by the consequence of a failure, in particular whether the failure is classified as dangerous or safe. To explore the possibilities for automated classification of failure events, the paper first presents a taxonomy needed to annotate failure notifications. The suggested safety critical equipment groupings, and tailor-made taxonomies were developed in close collaboration with the industry. An important basis for the taxonomies has been the International Standards Organization 14224 standard for collection and exchange of reliability and maintenance data for equipment in the petroleum, petrochemical, and natural gas industries. The taxonomy for failure modes was tested in an industry workshop where a group of experts used an annotation tool on a dataset of notifications for shutdown valves. The annotation approach, results, and agreements between the experts are presented. Several issues were identified during the study, providing a better understanding of the challenges, possibilities for improvement and pitfalls that should be avoided when annotating larger datasets.

1. INTRODUCTION

Previous research has investigated how informal technical language written in maintenance work orders can be mapped to a computable format with the goal of unlocking maintenance knowledge and support data analysis and maintenance decisions for different equipment groups. An example of recent work in this area is Sexton, Hodkiewicz, Brundage, and Smoker (2018). The technical language found in maintenance work orders is unstructured and informal, and often contains abbreviations or domain specific terms that makes it difficult to employ conventional natural language methods as these are usually designed for non-technical use-cases. To overcome this issue, researchers have used Technical Language Processing (TLP), which combines Natural Language Processing (NLP) techniques with technical domain specific dictionaries and the identification of meaningful domain-relevant relationships in the data (Gao, Woods, Liu, French & Hodkiewicz, 2020, Brundage, Sexton, Hodkiewicz, Dima & Lukens, 2021).

Annotation or tagging of text in maintenance notifications and work orders performed by experts, is a technique to prepare unstructured text for analysis, classification and decision making. The tagged data can be used to identify information and relationships that have previously been hidden or inaccessible, and to build a dictionary of entities and aliases that can be used to tag a larger dataset automatically. Since the information gained from analyzing the tagged data can be used in important maintenance decisions, it is crucial that these tags are accurate and reliable. From experience, we know that there is inconsistency in the way failures are reported and interpreted. Also, the datasets are often large, and it would be unrealistic to assume that a single person can annotate the whole corpus. Therefore, the tagging process must in practice be done by multiple experts, requiring that a sufficient agreement between the experts is obtained beforehand. These issues are also discussed in the work by Hastings, Sexton, Brundage, and Hodkiewicz (2019). The work presented in this paper aims to investigate how TLP techniques can be applied to facilitate more efficient and automated classification of failure events for safety equipment on petroleum facilities. Examples of typical equipment groups are gas detectors, level transmitters, logic solvers, and shutdown valves. The latter is used as example equipment in the present paper.

Requirements to be followed up during operation of safety critical equipment on a petroleum facility are specified by the Petroleum Safety Authority (PSA) Norway regulations and international standards such as IEC 61508 and IEC 61511 (International Electrotechnical Commission, 2010, 2016), in internal company governing documents as well as facility specific requirements. The safety integrity level (SIL) and the corresponding probability of failure on demand (PFD) are specific reliability requirements that should be verified regularly during operation. The PFD of a component is a function of the dangerous undetected (DU) failure rate and the proof test interval of the component. A DU failure is a failure that prevents a component from performing its safety functions and where the failure is not revealed immediately after occurrence, but rather during proof tests or on demand. An example of a DU failure is a shutdown valve that fails to close upon a proof test.

A fundamental concept in both IEC 61508 and IEC 61511 is the notion of risk reduction; the higher the risk reduction is required, the higher the SIL. It is therefore important to apply *realistic* failure data in the design calculations, since too optimistic failure rates may suggest a higher risk reduction than what is obtainable in operation (IEC, 2010, 2016). This is emphasized in the second edition of IEC 61511-1 (sub clause 11.9.3) (IEC, 2016) which states that the applied reliability data shall be credible, traceable, documented and justified and shall be based on field feedback from similar devices used in a similar operating environment. To fulfil these requirements, high quality reporting and classification of failures and failure notifications are crucial. However, the collection and classification of failure notifications and identification of DU failures to obtain realistic failure rates are rather time consuming, and the notification quality varies. Experience has shown that essential information to classify notifications may be missing and/or that it is necessary to investigate other sources (such as event logs or condition monitoring systems and discussing with technicians and equipment experts). An example is the close to 30.000 notifications manually classified to obtain the failure rates presented in the PDS data handbook (Ottermo, Hauge & Håbrekke, 2021). Based on this work we estimate that classifying 500 notifications takes three to four days, and involves between four and six persons, which amounts to a total of approximately four man-labor years for classifying all notifications.

Several domains need to be investigated to enable automated and consistent classification of failure notifications. As a starting point, we have tested the annotation approach and agreement behavior on a group of experts that were asked to classify and annotate failure modes for shutdown valves. A suitable web-based annotation tool was used in a web meeting with the experts. The annotation tool, Redcoat, is described in more detail in section 3.3.

1.1. Paper outline

The outline of the paper is as follows: Section 2 describes how failure reporting and classification for some relevant parameters are performed today and discusses potential for automation. Section 3 describes the established taxonomy for some of the parameters for the selected equipment group, the dataset, and the annotation tool. Section 4 describes the annotation of a set of notifications. The results are presented in section 5. Finally, the discussion and conclusion are given in sections 6 and 7, respectively.

2. FAILURE REPORTING AND CLASSIFICATION

The collection of maintenance and failure data is often subject to concerns about the adequacy, quality, and uncertainty of the data. An important starting point for addressing these concerns is to ensure that failures are reported and classified in a *consistent* way.

Figure 1 illustrates a typical workflow for failure event reporting, classification, and further data analysis, with focus on 'Failure reporting' and 'Failure classification' (indicated with the dotted, red circle). Three of the most important parameters for reporting and classification of failure notifications, are detection method, failure mode, and failure cause, and these are described in more detail below. Today, these parameters are manually registered in the notifications, and in the maintenance systems of the oil and gas operators in Norway, they can usually be selected from a list of alternatives mainly based on ISO 14224 (International Standards Organization, 2016). ISO 14224 provides a basis for collection of reliability data for oil and

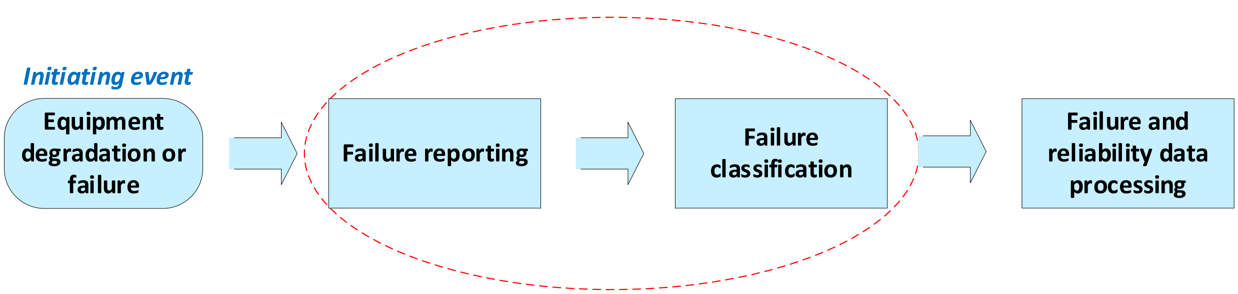

Figure 1. Workflow for failure event, reporting, classification, and further data analysis

gas equipment including equipment groups and lists of possible detection methods, failure modes, failure mechanisms, and failure causes for the equipment groups. The technicians reporting failure notifications are not always familiar with the listed alternatives and the meaning of all possible categories, particularly for the failure mode. As a result, the 'other' or 'unknown' categories are often selected for the failure mode. An internal study recently performed over a six-month period for a Norwegian offshore facility showed that for more than 50% of the notifications, the failure mode was classified as either 'other' or 'unknown'. Hence, the most relevant information about failure mode, as well as detection method and failure cause, is often to be found (or most likely to be trusted) in the free text field(s) in the notifications.

As part of the failure classification, each notification is classified as either DU, dangerous detected (DD), safe or non-critical. Both DU and DD failures are *dangerous* failures preventing the component from performing its safety function, while *safe* failures maintain the component in a safe state. Non-critical failures may be degraded failures that have not (yet) developed into a (safe or dangerous) failure. *Detected* failures are revealed 'immediately' after occurrence, for instance a frozen level transmitter that is alarmed in the control room or a line gas detector that provides an alarm due to blocked beam. *Undetected* failures are not revealed immediately and may be latent until the component is tested or demanded. Assuming that the DD failures are corrected within short time, only the DU failures contribute to significant unavailability of the safety function and are important to identify.

The **detection method** characterizes how the failure was discovered and determines if the failure is classified as *detected* (i.e., revealed immediately) or *undetected* (i.e., latent until test or demand). To evaluate the quality and effectiveness of the maintenance program, it is also important to know whether the failure has been reported automatically (by self-diagnostics) or by personnel, in relation to planned and regular activities (such as preventive maintenance and regular testing), or in relation to irregular activities such as demands, corrective maintenance or random observations.

A **failure mode** characterizes how a failure is brought to our attention, i.e., in what way it is observed that the function has been fully or partially lost. Based on the failure mode, the equipment group, and its safety function, it can be determined if the failure is *dangerous* or *safe.* As for detection method, this is considered key information and most maintenance systems provide a pre-defined list of failure modes to select from. A valve's safety function will typically be to open or close on demand, and to keep tight in the closed position. Failure to close or open when demanded, an internal leakage in closed position, and untimely (spurious) activation are all examples of failure modes. Correct reporting of failure modes is essential for determining the severity of the failure.

A **failure cause** or root cause characterizes the circumstances or conditions that lead to a failure, or the initiating event in the sequence of events leading to a failure. The failure cause (or root cause) is usually not known when the failure notification is registered and may be revealed later during troubleshooting and repair of the component. A failure cause or root cause analysis may give additional information about the probability of the failure to re-occur and the need for mitigating measures to prevent this from happening. However, such analysis is time consuming and therefore only rarely performed.

2.1. Potential for automation

'Automated process for monitoring of safety instrumented systems' (APOS) is an ongoing research project that aims to help industry partners develop new solutions for collecting, analyzing, and sharing failure data for safety systems during design and operation of a petroleum facility. The project develops knowledge and specifications that simplify and automate design and operation of safety systems. One of the main activities in the APOS project is to explore how machine learning techniques can be applied to interpret information added as free text and automatically identify and classify failures. This information can possibly be combined with information from the safety and automation systems (SAS), information management systems (IMS) and condition monitoring systems to improve the quality of the automated classification, and then be utilized for efficient decision making, such as updating of proof test intervals and identification of risk reducing measures.

To simplify failure reporting and classification, algorithms that can reduce parameter choices are suggested. One possibility is to apply available event information from SAS and IMS. Another possibility is a further limitation of failure modes based on the equipment group under consideration reducing the number of alternatives at the same time as a classification into dangerous and safe is possible. Also, the categories 'other' and 'unknown' are suggested omitted. Taxonomies with two or three levels are suggested both for detection method and failure mode. The taxonomies are described in more detail in section 3.

Having determined the failure mode, it will then be possible to distinguish between DU and DD failures based on the detection method: The detection method tells if the failure is detected or undetected and the (equipment group specific) failure mode tells if the failure is dangerous or not, see Figure 2.

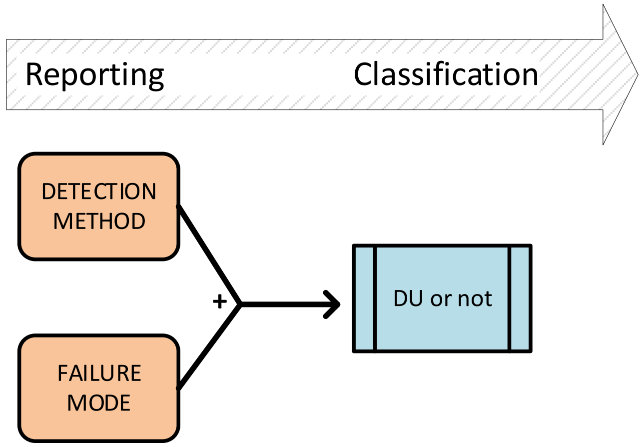

Figure 2. Automatic classification of failures based on detection method and failure mode

3. TAXONOMY, DATA, AND TOOL

3.1. Taxonomies for annotation

One of the deliveries of the APOS project is to propose a simplified and standardized taxonomy for safety equipment grouping, detection methods, failure modes and failure causes based on common practices from operators in the Norwegian oil and gas industry. Based on this work, the fundamental taxonomy needed during annotation of safety critical equipment was developed (Hauge, Håbrekke & Lundteigen, 2020).

The suggested detection method and failure cause taxonomies are (for the time being) common for all equipment groups. The suggested detection method taxonomy (modified from ISO 14224 (ISO, 2016) for safety critical equipment has two levels and is shown in Table 1. The corresponding ISO 14224 categories (from Table B.4 in the standard) are listed in the rightmost column of Table 1. Note that five detection method categories are suggested as compared to eleven (including 'other') in ISO 14224.

The suggested failure mode taxonomy is equipment group specific, both since different equipment fail in different ways and since the criticality of a failure mode may be dependent on the equipment's functionality. Also, within equipment groups there may be some subgroups of components having different functional requirements. For instance, some shutdown valves have leakage requirements and others do not. Hence, the failure mode leakage in closed position (LCP) for a shutdown valve with leakage requirement is defined as 'dangerous' but for a shutdown valves *without* leakage requirement, the failure is defined as degraded' or 'non-critical'. The suggested failure mode taxonomy for shutdown valves has two levels and is shown in Table 2.

Table 1. Detection method taxonomy

D0	Detection method (level 1)	Detection method (level 2)	Corresponding ISO 14224 categories
Undetected	1. Scheduled activities	1.1 Functional test	02 Functional testing
		1.2 Other periodic maintenance (PM) activity	01 PM 03 Inspection 04 Periodic condition monitoring
	2. Unscheduled activities and events	2.1 Demand	07 Production interference 10 On demand
		2.2 Casual observation	05 Pressure testing 08 Causal observation 09 Corrective maintenance
Detected	3. Alarmed upon occurrence 2)	3.1 Diagnosed / immediately detected event	06 Continuous condition monitoring

Table 2. Failure mode taxonomy for shutdown valves

F0	Failure mode (level 1)	Failure mode (level 2)
Dangerous failures	Dangerous failure	Fail to close (FTC)
		Delayed operation (DOP) * 1)
		Leakage in closed position (LCP) * 2)
Safe failures	Safe failure	Fail to open (FTO)
	Spurious failure	Spurious operation (SPO)
Non-critical (NONC) failures	Degraded failure	DOP * 1)
		LCP * 2)
		Structural deficiency (STD)
	No-effect failure	Noise (NOI)
		Abnormal instrument reading (AIR)
		Minor in-service problems (SER)
Non-functional-safety failures	Loss of containment	External leakage – utility medium (ELU)
		External leakage – process medium (ELP)
	Loss of explosion (EX) protection	Loss of EX protection (LEX)

1) Only relevant if response time requirement given.

2) Only relevant if internal leakage requirement given.

* the same failure mode on level 2 is relevant for more than one failure mode on level 1 (for instance. an internal valve leakage above versus below the specified acceptance criterion).

3.2. Dataset

The dataset applied in this study (see section 4) was an extraction of 80 notifications from three different oil and gas facilities in Norway for emergency shutdown valves. The text fields were mainly based on the short text field of the notifications. For those notifications with limited text in the short text field (1–2 sentences), the long text field was added. Normally the long text field comprises relevant information regarding failure mode, but often also additional information about possible failure causes, corrective maintenance work, suggested mitigating measures, etc. All data were anonymized, meaning that tag number and names of persons, facilities, etc. were removed.

3.3. Annotation tool

To annotate the dataset, an annotation tool was selected based on the following criteria:

- Suitable tool with respect to technical language processing.
- User friendly, i.e., easy to annotate – particularly for persons not familiar with the tool.
- Possibility to establish a taxonomy or hierarchy of entities without any restrictions on numbers and levels, and possibility to adjust the hierarchy during the annotation process.
- Tailored for collaborative annotation.
- Optimized for short texts.

The web-based annotation tool Redcoat (Stewart, Liu & Cardell-Oliver, 2019) was found to fulfil the above criteria and was selected for our study. Particularly, the tool is user friendly and has the possibility to involve several participants by simple user registration (and with no need for each participant to download the tool). Redcoat features the possibility to establish separate projects, each with a defined hierarchy of entities, and an uploaded dataset of short text fields. After performing the annotation, each project participant can view the results from the annotation performed by all participants.

4. STUDY - DESCRIPTION

The study presented was limited to shutdown valves and failure mode annotation and comprised the following two steps:

1. A project was established in Redcoat, and the taxonomy with the failure modes for shutdown valves (Table 2) were defined in the project. The Redcoat tool, the taxonomy, and the dataset were tested by a limited group of experts.
2. The results from the test group were used to refine the annotation strategy, the taxonomy, and the selection of the notifications in the dataset. A larger group of persons, having different positions in the petroleum industry, but all working with safety systems, were gathered in a workshop, and engaged in the entity typing of the notifications.

4.1. Expert group testing

For the expert group testing, two persons familiar with the annotation tool and three persons unfamiliar with the tool were included. The participants were denoted as experts, in the sense that they had considerable experience with classification of notifications of safety critical equipment. The participants unfamiliar with the annotation tool were given a short demo before starting the annotation.

Even though all experts had considerable experience, a few issues were encountered:

1. For the taxonomies with two or three levels, only level 1 is compulsory. Therefore, both level 1 and level 2 from Table 2 were included in the entity hierarchy. When doing the actual testing, this turned out to be too complicated, especially for experts not familiar with the hierarchy or the tool. Therefore, only level 2, i.e., the most detailed level, of the failure modes was included in the annotation work.
2. Since notifications in the Norwegian oil and gas industry are often written in languages other than English (often a mixture of Norwegian, Danish, and Swedish, in addition to English), words containing special characters are split when imported to Redcoat (space introduced before and after special character). For instance, in the Norwegian language the characters 'æ', 'ø' and 'å' splits the words. This required a preprocessing of the data by replacing 'æ' with 'ae', 'ø' with 'oe' and 'å' with 'aa'.
3. Due to restricted time and limited experience with annotation and TLP, the number of notifications was reduced from 80 to 25.

4.2. Workshop testing

Prior to the workshop testing, a simple user manual was developed for the participants, including how to login to Redcoat, how to accept and open a project, and how to perform annotation. None of the participants had any experience with TLP prior to the workshop. In the beginning of the workshop a brief introduction to TLP, the tool, and the annotation task was given. Then, a total of 20 participants – all familiar with safety critical equipment – performed individual annotations of the notifications with respect to failure modes. All participants were asked to annotate the same 25 notifications.

In addition to the participants that did individual annotation, a co-operative annotation session in a larger group of system experts was facilitated. About ten of the notifications were reviewed and discussed. The discussion among the participants during the annotation session and the subsequent plenum discussion, revealed several interesting issues regarding the annotation.

5. RESULTS FROM WORKSHOP

The agreement for each notification varied between 13% and 68%. An average agreement of 41% was obtained. The main causes of disagreement were:

1. <u>Number of words included:</u> Several missing agreements arose because the participants included various number of words to describe the failure mode (see Figure 3). From the figure we see that there is agreement concerning the failure mode ('delayed operation'). However, some annotators have included the entire sentence, some have avoided typical fillers (such as 'paa'), and others have been restricted to keywords (such as 'closing time' (Norwegian: 'stengetid')). It should be noted that limited instructions on how many words to include were given in advance of the actual annotation.

2. <u>Number of failure modes chosen:</u> In many cases, there were two or more opinions concerning which failure mode to choose due to limited information in the given text. Prior to the workshop, the participants were shown how to select several failure modes. It was informed that when uncertainties arose, *all possible* failure modes should be selected. For instance, if the notification says: 'valve does not move', it is not possible to know whether the failure mode is 'fail to close' or 'fail to open'. However, from the results it seems that many chose only the most probable failure mode(s) or only the first failure mode the annotator associated with the text. Also, some annotators may have selected only the *worst case*, such as 'fail to close', which is most critical for a shutdown valve (see Figure 4).

3. <u>Different failure modes chosen:</u> Due to subjective understanding of both the text describing the failure and the failure mode, the annotators selected various failure modes, see Figure 5. For the example 'alarm on closing time' (Norwegian: 'alarm paa stengetid'), the annotators have selected various failure modes depending on their understanding of the failure leading to this alarm: 1) The shutdown valve not closing 100% and thereby not giving a closing alarm, 2) The valve closing 100% but not within the response time requirement (also giving an alarm), or 3) The valve has closed spuriously, and this spurious closure is revealed by an alarm.

   Another example of potential for individual understanding of the text was related to distinguishing between failure mode and failure cause. While some annotators always understood the notification text as the failure mode, other annotators understood the same failure mode as a failure cause leading to another possible failure mode. For instance, text including 'hydraulic leakage' was by some annotators interpreted as external leakage of utility medium. However, some pointed out that this is a possible failure cause that can lead to a 'fail to close'.

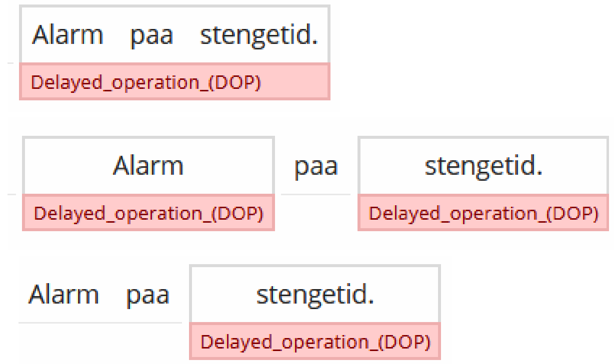

Figure 3. Missing agreement between annotations due to number of words included
('Alarm paa stengetid' = 'Alarm on closing time')

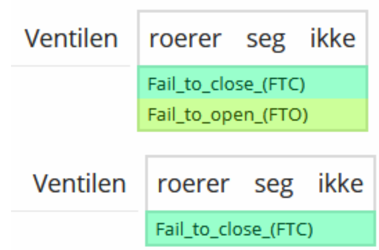

Figure 4. Missing agreement between number of failure modes included in the annotation
('Venitlen roerer seg ikke' = 'Valve does not move')

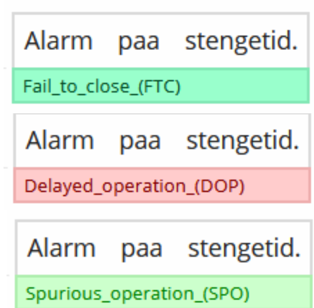

Figure 5. Missing agreement between selected failure modes

The entity frequencies of the failure modes are illustrated in Figure 6. The frequency of a failure mode equals the total number of annotations in the 25 notifications among all 20 participants. All 12 failure modes in level 2 of the established taxonomy (see Table 2) have been used at least once. The most frequent failure modes are external leakage of utility medium (ELU), fail to open (FTO), fail to close

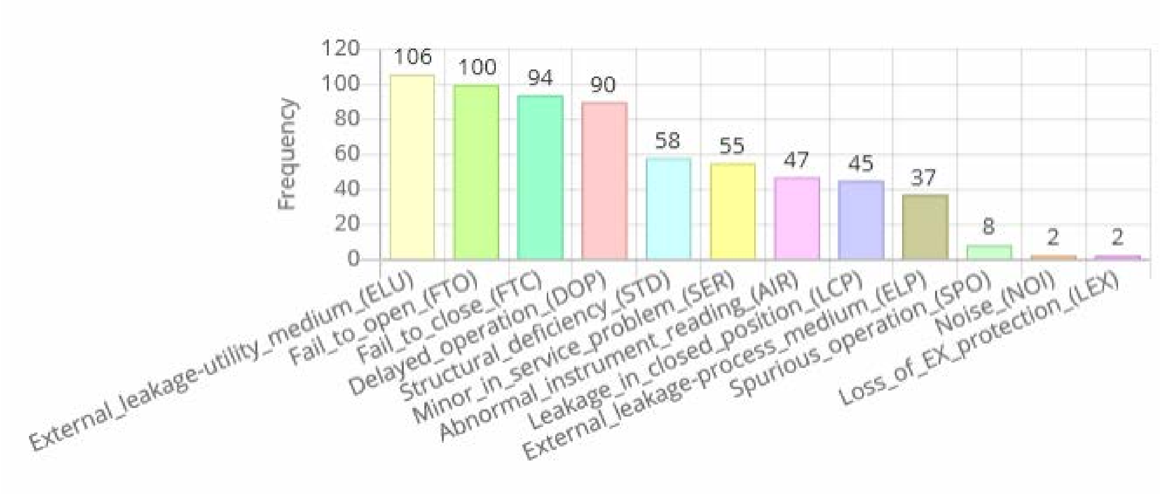

Figure 6. Total entity frequencies

(FTC), and delayed operation (DOP). This result may be interpreted in several ways:

1. ELU, FTO, FTC, and DOP are the most common failure modes for shutdown valve. Of course, this assumption needs to be further substantiated based on a larger dataset than only 25 notifications.
2. ELU, FTO, FTC, and DOP occur most frequently since these failure modes are often annotated together with other failure modes, for instance, FTO and FTC have often been selected for the same text due to uncertainty, ref. above.
3. ELU, FTO, FTC, and DOP are the most familiar failure modes to the annotators, either with respect to criticality (as FTC, DOP, and FTO are all dangerous failures for shutdown or blowdown valves) or with respect to occurrence (for example ELU is a well-known problem for valves).

6. DISCUSSION AND FURTHER WORK

6.1. Discussion

As pointed out in the result section, the dataset used is too small for statistical analysis. Despite this, several issues were encountered that can be used actively to improve and prepare for the annotation of bigger datasets. Some needs and improvements both regarding the annotation and the failure reporting and classification are also discussed.

In the study described, the only available information for the participants, besides the text, was the background information about the equipment group and its safety function (i.e., shutdown valves). Further information, such as response time requirements, leakage requirements, process application, valve design, etc., that may be relevant when performing failure mode classification was missing for most of the notifications included in the study. Such context information that can be helpful for choosing the correct failure mode is often available from other fields in the notifications, from the associated work orders or from SAS, IMS and condition monitoring systems. Still, the machines and programs that should learn how to annotate the dataset, do not necessarily have this context information, even if combined with other fields in the notification. Hence the annotations should, ideally, be context independent.

Some of the workshop participants from the oil and gas companies pointed to the need for increasing the awareness among the technicians on how to improve the text they write in notifications and work orders. It is often difficult to describe the failure precisely and descriptively due to lack of detailed knowledge, and failures often need to be further investigated to determine the cause. Also, the technicians may have limited training in classifying failures with respect to failure mode. Therefore, it was suggested to use such annotation tools actively in training, which could result in both improved notification texts and better understanding of the systems.

One of the participants also suggested to add some intelligence into the free text field, i.e., to provide text suggestions for the operator while writing the notifications. Relevant suggestions to detection methods, failure modes and failure causes could for instance be highlighted for the technicians. Another possibility is that relevant questions appear based on the text already entered, such as 'does this imply that the valve failed to close or failed to open?'. Also, additional intelligence could give the possibility to reveal and correct misspellings, abbreviations and terms meaning the same thing (such as not respond / no response) online as the free text is entered. Additional to the mentioned challenge that various languages and terms are used among Norwegian oil and gas employees, it was also pointed out that the way technicians write and communicate is developing. Newly educated technicians may have another vocabulary and use other abbreviations today as compared to 10–20 years ago, and this can result in a huge number of aliases for one single entity.

Finally, it needs to be pointed out that the workshop and annotation exercise described in this paper, for simplicity, only focused on selecting the correct failure mode and annotating this against one or more words in the free text description. In a real situation it will be more relevant to annotate against more entity types, thus reducing some of the uncertainty discussed in section 5.1, for instance with respect to how many words to select.

6.2. Further work

As discussed above, there is a need to annotate more data. Participants in the workshop have therefore been invited to perform additional annotations of a larger dataset comprising both shutdown valves and point gas detectors. The possibility for annotating against more entity types will be included in the project. Based on this, a dictionary can be established that will enable us to evaluate the performance of automated classification of our database of several thousand manually classified notifications collected during the last decade.

7. CONCLUSIONS

The aim of this work has been to investigate how technical language processing can be utilized to classify failure events for safety critical equipment. Today, the failure classification is a manual and tedious process prone to errors and subjective bias. An initial annotation study has been conducted to form the basis for a larger annotation project. Several issues have been identified during the study, providing a better understanding of the challenges and the pitfalls that should be avoided when annotating larger datasets, for instance to be more specific about how many failure modes to select for the same entity. Some new possibilities, beyond making the classification process more efficient and consistent, were also identified. One example is to use the annotation tool for training of operators both with respect to writing more consistent notifications as well as getting a better understanding of problems they might be facing in the field. Also, additional event information from SAS and IMS could be used to provide online assistance when reporting failures.

ACKNOWLEDGEMENT

The paper presents results from APOS, a joint industry project on automized process for follow-up of safety instrumented systems. The project is supported by the Norwegian Research Council (Project no. 295902) and 11 industry partners representing oil companies, engineering, consultants, and vendors of control and safety systems.

NOMENCLATURE

APOS	Automated process for monitoring of safety instrumented systems'
DD	Dangerous detected
DU	Dangerous undetected
EX	Explosion
IEC	International Electrotechnical Commission
IMS	Information management system
ISO	International Standards Organization
NLP	Natural language processing
PFD	Probability of failure on demand
PM	Preventive maintenance
PSA	Petroleum safety authority Norway
SAS	Safety automation system
SIL	Safety integrity level
TLP	Technical language processing

REFERENCES

Brundage, M., Sexton, T., Hodkiewicz, M., Dima, A., & Lukens, S. (2021). Technical Language Processing: Unlocking Maintenance Knowledge, *Manufacturing Letters*, Volume 27, January 2021, Pages 42-46.

Gao, Y., Woods, C., Liu, W., French, T., and Hodkiewicz, M. (2020), Pipeline for machine reading of unstructured maintenance work order records, *Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference (ESREL2020 PSAM15).*

Hastings, E. M., Sexton, T., Brundage, M.P., & Hodkiewicz, M. 2019. *2019 Annual Conference of the Prognostics and Health Management Society.*

Hauge, S., Håbrekke, S., & Lundteigen, M. A. (2020). Standardised failure reporting and classification of failures of safety instrumented systems. *Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference (ESREL2020 PSAM15).*

International Electrotechnical Commission (IEC) (2010). IEC 61508. Functional safety of electrical/electronic/programmable electronic safety-related systems. International Electrotechnical Committee.

International Electrotechnical Commission (2016). IEC 61511. Functional safety – safety instrumented systems for the process industry sector. International Electrotechnical Committee.

International Standards Organization (ISO) (2016). ISO 14224. Petroleum, petrochemical, and natural gas industries – Collection and exchange of reliability and maintenance data for equipment. ISO 14224:2016, Corrected version 2016-10-01.

Ottermo, M., Hauge, S., & Håbrekke S. (2021) Reliability *Data for Safety Equipment, PDS Data Handbook,* Trondheim, SINTEF Byggforsk.

Sexton, T., Hodkiewicz, M., Brundage, M. P., & Smoker, T. (2018). Benchmarking for keyword extraction methodologies in maintenance work orders. *In Proceedings of the Annual Conference of the PHM Society (Vol. 10, No. 1).*

Stewart, M., Liu, W., & Cardell-Oliver, R. (2019, November). Redcoat: A collaborative annotation tool for hierarchical entity typing. *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstration.*