ACTIVE AND KNOWLEDGE-BASED PROCESS SAFETY INCIDENT RETRIEVAL SYSTEM

A Thesis

by

SARA SHAMMI KHAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2010

Major Subject: Safety Engineering

Active and Knowledge-based Process Safety Incident Retrieval System

ACTIVE AND KNOWLEDGE-BASED PROCESS SAFETY INCIDENT

RETRIEVAL SYSTEM

A Thesis

by

SARA SHAMMI KHAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	M. Sam Mannan
Committee Members,	Marietta J. Tretter
	James Caverlee
Head of Department,	Michael Pishko

August 2010

Major Subject: Safety Engineering

ABSTRACT

Active and Knowledge-based Process Safety Incident Retrieval System. (August 2010)
Sara Shammi Khan, B.S., Texas A&M University
Chair of Advisory Committee: Dr. M. Sam Mannan

The sustainability and continued development of the chemical industry is to a large extent dependent on learning from past incidents. The failure to learn from past mistakes is rather not deliberate but due to unawareness of the situation. Incident databases are excellent resources to learn from past mistakes; however, in order to be effective, incident databases need to be functional in disseminating the lessons learned to users. Therefore, this research is dedicated to improving user accessibility of incident databases. The objective of this research is twofold. The first objective is improving accessibility of the database system by allowing the option of word search as well as folder search for the users. This will satisfy research needs of users who are aware of the hazards at hand and need to access the relevant information. The second objective is to activate the database via integration of the database with an operational software. This will benefit research needs of users who are unaware of the existing hazards.

Literature review and text mining of Major Accident Reporting System (MARS) database short reports are employed to develop an initial taxonomy, which is then refined and modified by expert review. The incident reports in MARS database is classified to the right folders in the taxonomy and implemented in a database system based on Microsoft Excel, where the users can retrieve information using folder search as well as word search option via a user friendly interface.

A program coded in JAVA is prepared for integrating the incident database with a Management of Change (MOC) software prototype. A collection of keywords on hazardous substances and equipment is prepared. If the keywords exist in the MOC

interface, they will be highlighted, and with the click of a button, will return up to ten relevant incident reports. Using an active and knowledge-based system, people can learn from incidents and near-misses and will be more active to reduce the frequency of recurring incidents.

DEDICATION

To

Ammu & Abbu, through whose unconditional love and sacrifices have I been able to achieve all that I have in life;

My beloved husband, *Zami*, whose endless love and pampering has spoiled me beyond measures;

My siblings: *Shaila*, *Shakil*, *Shoma*, *Masoom*, *Anika*, *& Rumi*, whose love and constant support have kept me going.

ACKNOWLEDGEMENTS

I would like to thank my academic advisor, Dr. Sam Mannan, for his endless encouragement and support throughout my graduate study. A person with great principle and concerns for his students, he will remain a role model in my future. I would like to express my gratitude to my committee members, Dr. Marietta Tretter and Dr. James Caverlee, for their teaching and guidance inside and outside the classroom.

It is an honor for me to thank Dr. Trevor Kletz, Dr. Jack Chosnek, Dr. Hans Pasman, and many other visiting scholars at MKOPSC, without whose expert advice and guidance, this thesis would not have been possible. I want to extend my gratitude to Dr. Christian Delvosalle from Faculté Polytechnique de Mons, for giving me permission to use the ARAMIS project data and Mr. Lorenzo from the European Commission-Joint Research Center, for allowing me to use the MARS database for my research. I also would like to acknowledge Dr. Adrian Sepeda for sharing information about the CCPS Process Safety Incident Database.

I am grateful to Pravin Prakash, who has gone beyond his job responsibilities to help me finish this thesis, and to Karthik Ramachandran, who has provided the programming support. I appreciate the staff, team leaders, and my colleagues at Mary Kay O’Connor Process Safety Center, especially Suhani, Katherine, Mahdiyati, Yuan, Jiejia, Alberto, Seungho, Victor, Dr. William Rogers, Dr. Dedy Ng, Valerie Green, and Mary Cass for all their help in making my experience at the center a memorable one. Thanks also goes to Towanna Arnold for her help with all my paperwork.

Finally, thanks to my mother and father for their encouragement, and to my husband for his love and support.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

1. INTRODUCTION

We cannot wait to learn from our own mistakes, thus, incident databases serve as invaluable resources to learn from mistakes made by others in the past, which are often accompanied with appalling consequences. Incident databases provide insights on what can go wrong with various processes, equipment, and operators and how devastating the outcome can be from a rather trivial hazard, which could go unseen during a Process Hazard Analysis (PHA).

In order to be effectively useful incident databases should be comprehensive, functional, and should offer core learning from the past mistakes. A database is comprehensive, when it has a complete and accurate inventory of all different types of incidents. Core learning is gained whenever a detailed incident investigation is done to capture the root causes of the incidents and the results are well documented. Finally, for it to be functional, a database should be well structured and should be easily accessible to the users.

The proposed research is dedicated to improving the functionality of incident databases by:

1. Giving structure to the database by means of an intelligent incident taxonomy;
2. Providing a ‘word search’ as well as a ‘folder search’ system for users, thus enhancing the information accessibility; and
3. Integrating incident database to an operational software thereby converting it to an ‘active’ system, which can direct users to the incident information of interest.

____
This thesis follows the style of the journal of *Industrial and Engineering Chemistry Research.*

1.1 Motivation and Objective

The sustainability and continued development of the chemical industry is to a large extent dependent on learning from past incidents, which is evident by the repeated mistakes in design and operation leading to the recurring incidents.[1, 2] As any other industry, chemical industry strives to eliminate hazards and reduce incidents, since incidents are often tagged along with enormous financial losses and human casualties, which also diminish reputation of the industry. Consequently the failure to learn from past mistakes has significant consequences.

In the context of potential hazards that exist in chemical process industry, there are two categories of unawareness or lack of knowledge: the first is 'knowing that you don't know' the second is 'not knowing that you don't know'. [3] Both of these categories incur hazard, however the second is more undesirable, because unawareness of the potential danger eliminates the opportunity to research, identify and react appropriately to reduce or eradicate the hazard.[3] Therefore in improving process safety data systems, the objective of this research is twofold:

1. The first objective is to improve the traditional database search system by structuring the database through classification of the incident reports in an intelligent taxonomy. The taxonomy stored in a database and the incident summaries classified and assigned to the respective nodes of the taxonomy tree, will allow folder searching to enhance accessibility for the users. The resulting knowledge-base would benefit the research needs of users, who are already aware of the potential hazards of the process setting.

2. The second objective is to transform the incident knowledgebase to an 'active' system, where the passive users will be alerted of unknown hazards by the system. This transformation will be done by integrating the database with an operational software – specifically a Management of Change (MOC) software prototype for this research, as shown in Figure 1.

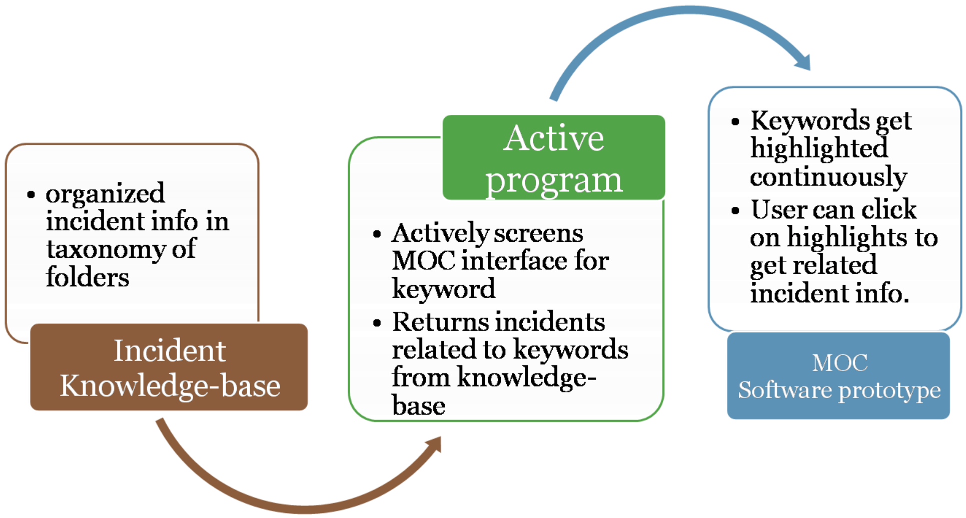

Figure 1. Diagram showing the active functioning of the database through integration of the standalone incident database with MOC software prototype

Having an incident database with its unique guided search system and modern information retrieval techniques and linking it with operational software would provide a powerful tool and would improve safety and hence reduce costs.

2. BACKGROUND

2.1 Existing Databases

The Center for Chemical Process Safety (CCPS) have developed an incident database to share incident information among members and enhance lessons learned. This unique database called Process Safety Incident Database (PSID) has undergone numerous upgrades and now offers a web-based application, which enables accessibility by any PSID participant at anytime, anywhere.[3] Among other features that enhances functionality, such as data quality, potential of constant growth, security of data and confidentiality of data contributor, database management, and web accessibility, PSID also provides a conventional well-developed query system. The query system includes a few built in features such as the two dimensional (2-D) cross plot, whereby the user can plot any two data fields against each other and view statistical relationships.[3] This feature is useful to learn the current industry trends and to efficiently allocate resources in order to minimize or prevent losses. In addition, PSID offers pre-programmed lists of topics that are very useful for inexperienced users, who can utilize this list to search for topics of interest. This list can also be customized by experienced users, who would like to keep their common search topics saved for quick access.

The database search system developed in this research is similar to that of CCPS-PSID in the way that it offers folder search and word search. However, the PSID has a single level of folders (keyword list) in the hierarchy, after the broad divisions of *type of operation, chemical involved, phase of operation, initiating event, incident type, contributing factors, equipment type,* and *incident cost.* In this research, the folders are arranged in a hierarchy up to a maximum of four levels. Moreover, word search within a folder has also been included for the search system, which further improves searching capabilities.

Some of the existing incident databases that are open to public or accessible by Mary Kay O'Connor Process Safety Center (MKOPSC) members have been reviewed to analyze and compare the information retrieval options they provide. These databases are listed as follows:

Hazardous Substances Emergency Events Surveillance System (HSEES)

The HSEES database is established and maintained by the Agency for Toxic Substances and Disease Registry (ATSDR). Currently, there are fifteen participating state health departments that provide information on release events that fall under the definition of HSEES event. A HSEES event is any release of a hazardous substance that require decontamination and cleanup according to federal, state, or local law, or any threatened release that results in public action, such as evacuation.[4] HSEES however excludes any petroleum release event from reporting.

The available HSEES database is stored in Microsoft Access© format, where all the features for sorting data, customized queries to find data of a certain attribute, and word search capabilities that the database management software provides, is possible.

Environmental Protection Agency - Risk Management Program (EPA-RMP)

RMP 5-year Incident History database is maintained by the Environmental Protection Agency and covers incidents resulting in fatalities, injuries, and environmental or property damage involved with fixed facilities that store, use, or produce certain chemicals above a threshold quantity.[5] The data can be obtained in a CD in an application based on Microsoft Access.

National Response Center (NRC)

NRC serves as the single national point of contact for reporting all chemical, oil, biological, and radiological releases anywhere in the United States and its territories.[6] Manned by Coast Guard personnel NRC maintains 24 hours per day, 365 days per year telephone watch. The personnel enter incident reports obtained via telephone into the

Incident Reporting Information System (IRIS) and relay the report to the designated Federal On-Scene Coordinator (FOSC).[6] The collected data is made available to the general public online through the NRC website under the Freedom of Information Act (FOIA). NRC being the sole point of initial reporting for all type of discharges, the database is quite large and the data is not of the highest quality as the reporting has to be done within a short time frame, when people are dealing with the emergency and may not have all the information.

The NRC-IRIS online query system allows narrowing down search results relevant to specific criteria as listed below:

- NRC Report Number
- Call Type
- Incident Date
- Location County
- City
- State
- Suspected Responsible Company
- Type of Incident
- Medium Affected
- Material Involved

Thus, search result for the database can be refined by filling in the required criteria in the query form.

Occupational Safety and Health Administration (OSHA)

As an agency of the Department of Labor, OSHA's role is to assure the safety and health of American workers and prevent work-related injuries, illnesses, and fatalities by setting and enforcing standards.[7] OSHA also provides information on safety and health and educate, train and assist employers and workers to follow standards in order to keep a safe workplace. OSHA maintains an online database of incident investigation summaries. The summaries are in text form and can be queried through word search of the abstract or description of the incident. Besides word search they also allow searching by keywords and by Standard Industrial Classification (SIC) codes. The following is a list of the query fields in the OSHA online database:

- Description
- Abstract
- Keyword
- Display (fatality only)
- SIC
- OSHA Office
- Start Date
- End Date
- Inspection Number

OSHA incident investigation reports comprise mostly of personnel safety or ‘hard hat’ safety, since OSHA is concerned about all kinds of incident in the workplace. Keyword searching of the reports reveals that the keywords are not always efficiently tagged and query through keywords returns many irrelevant results.

Hazardous Material Incident Reporting System (HMIRS) and Integrated Pipeline Information System (IPIS)

HMIRS and IPIS are maintained by US DOT’s Pipeline and Hazardous Materials Safety Administration (PHMSA). DOT’s HMIRS database is the primary source of national incident data for hazardous materials transportation excluding pipelines. The data is collected from the carriers of dangerous goods through their reporting of dangerous goods incidents.[8] The database is downloadable from their website in comma separated value (CSV) format. HMIRS also maintains an extensive online query system to search incidents up to past ten years.[9] The query fields return categorical results and can be divided into six main categories:

- General incident information, which includes date and location of incident, mode and phase of transportation, carrier/reporter information
- Hazardous material name, including hazardous class and identification number.
- Packaging information
- Incident cause
- Incident consequence, including fatalities and injuries, fire, explosion, etc.
- Others, such as report number, container code, incident severity etc.

IPIS is the primary source for pipeline safety data reported by pipeline operators.[9] The database is downloadable from their website in CSV format and is divided into three sub databases: liquid pipeline; gas distribution; and gas transmission and gathering.

Major Accident Reporting System (MARS)

MARS database was established to manage major accident information submitted by the European Union (EU) member countries to the European Commission in agreement to the Seveso Directive.[10] The reporting system allows two different formats: 'short report' forms to be filled for immediate notification of incidents and 'full report' to be prepared after complete investigation of the incident. The short reports are available in an online database and are open to public. The reports are mainly text based with some categorical fields. The text fields are:

- Accident title
- Accident description
- Site description
- Installation/unit description
- Substance involved
- Causes of the accident
- Consequences
- Emergency response
- Lessons learned

The query system allows searching by incident date, legislation type, industry type, event type and full text search. The short reports comprise of only 10% of the number of the full accident reports and they are at times incomplete since these were reported when incident investigation has not been completed and the causes are unknown.[10] However the text description gives further insight on the causes and lessons learned and because of its concise format, it is ideal for text mining applications. Text mining of these reports can reveal unique words and relationship between these words, which can help in the incident taxonomy building process.

2.2 Existing Taxonomies

Chung and Jefferson[11] implemented an improved fuzzy search method that can exploit all available information in a database by returning not only the results that match a user defined query, but also those topics that have some relation to the query. To implement the fuzzy search method in an incident database, four incident classification hierarchies have been developed for the attributes operation, equipment, cause, and consequence. These classification hierarchies serving the sole purpose of supporting calculation of efficient Goodness of Fit (GOF) value (similarity score of the target data and the actual data) have been readjusted and extensively used wherever appropriate in building the taxonomy proposed for this research.

Anand[12] has proposed an equipment classification hierarchy similar to Chung and Jefferson’s classification. This taxonomy was used to generalize type of equipment into broad categories and used for data analysis to build a decision support system that would alert decision makers of potential hazards involving a chemical-equipment combination.

OSHA has adopted a new Globally Harmonized System (GHS)[13] approach to hazard classification, labels and safety data in order to increase the quality and consistency of information provided to employers, workers, and chemical users. OSHA GHS classifies chemicals according to their health and physical hazards. The taxonomy proposed for this research follows the OSHA GHS classification of chemicals but also includes classification by the state of the chemicals and other pertinent classes drawn from various sources.

Markowski[14] has developed classifications of several attributes in his work to improve and extend the application of Layer of Protection Analysis (LOPA) for the process industry. Method for incident scenario generation has been studied in his work, which resulted in the development of databases for selection of hazardous process, incident event, and layers of protection. These databases consist of classifications of different attributes namely, equipment, activity, hazardous substances, layers of protection, and causal factors, such as, process upsets, technical failure, human error, management oversights, and external events. The taxonomy as proposed in this research is partially based on Markowski's classifications.

Accidental Risk Assessment Methodology for Industries (ARAMIS) project, sponsored by the European Commission has defined a very comprehensive methodology for risk assessment. The method employs bow-tie diagrams as shown in Figure 2, to identify all major accidents assuming that the safety barriers are not in place or are inefficient.[15] Then the safety barriers are taken into account and evaluated to see if they are appropriate for the expected safety function. Finally, the evaluation of the probability of failure of the safety barriers enables to assess the incident frequency.

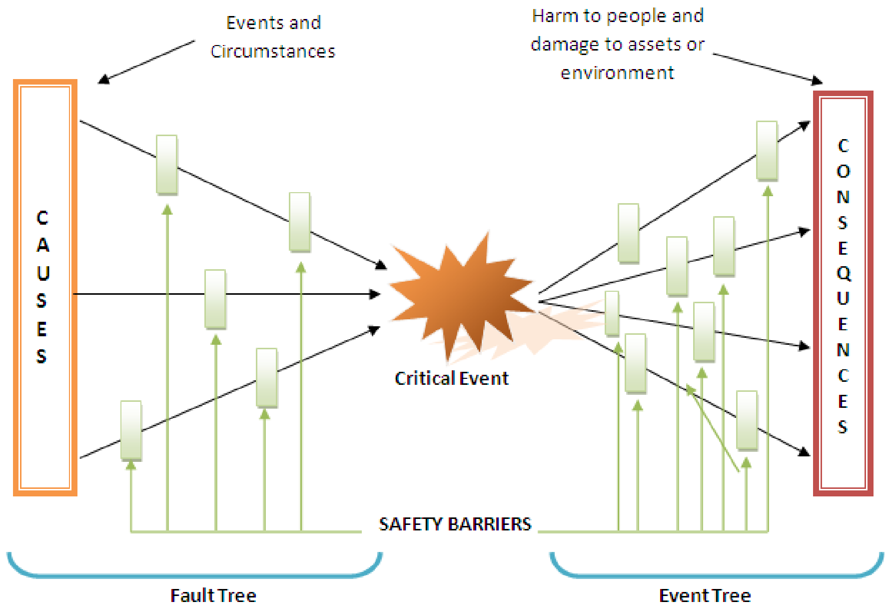

Figure 2. Complete bow-tie with identification of prevention or mitigation safety functions. (Adapted from Dianous and Fiévez,[15] and Hale et. al[16])

The ARAMIS project was in progress from 2001 to 2004 and prescribed some general classifications of equipment (16 categories), hazardous substances type, and critical event type (12 general categories). Also, 14 generic fault trees have been developed for different critical event types that can result from different combination of the defined equipment and hazardous material type. The fault trees start with the 'undesirable events' that can lead to the critical event (top event) via 'detailed direct causes', 'direct causes', and 'necessary and sufficient causes' stages, which depict the sequence of failure events. The event trees can be developed from the critical event onward with the aid of prescribed matrices, which leads up to the possible dangerous phenomena via the secondary and tertiary critical events.[17]

In this research, arranging the causes and consequences taxonomies according to the ARAMIS fault trees and event trees would have benefitted users who are using the incident knowledgebase for the purpose of scenario building. However, if the taxonomy is not structured in a hierarchy of categories, it will be complex and difficult for the casual users to understand. Hence, it would contradict the sole purpose of making the database more accessible. Therefore, to compromise, it is proposed to link some of the higher level categories (inner branches of the classification tree) in the main taxonomy to the ARAMIS fault trees for causes and event trees for consequences.

2.3 Integration of Incident Database with Operational and Design Software

Bond[18] has proposed the linking of an incident database to the design and operational software and demonstrated the advantages of this linkage to design software, to Hazard and Operability Studies (HAZOP) software and to risk assessment forms. It is very important to identify the hazards at an early stage or during the design phase of a process plant, since any modifications done later would be more costly. Therefore, it is important for the design engineer to be aware of all the past incidents involving the substances and equipment of the process to be designed. So if a database is integrated to design software, which actively alerts the design engineer of relevant past incidents with lessons learned information, it provides an opportunity to analyze all the hazards and incorporate the lessons learned in the final design. In risk assessment for an individual task, the first step involves identification of all the hazards. This step requires experienced people and also frequent look up in the incident database in order to identify the incident scenarios relevant to the task at hand. In Risk Assessment the combination of the incident database with the software provides a useful and user friendly system for identifying and assessing the risks. Process hazard analysis (PHA) is done before commissioning and also during operation (especially after a significant change has been made) of a process plant to identify the hazards that have been overlooked by designers and also those that exist for the unsuspecting operators. One of the common techniques used for PHA is HAZOP[1], which is a structured and comprehensive process of identifying hazards and requires a multidisciplinary team including experienced and knowledgeable individuals, process safety information (PSI) and past incidents. Therefore, integrating incident database to HAZOP software would reduce time involved in the HAZOP procedure and could also make the software more user friendly in identifying hazards.

Trevor Kletz[19] has proposed linkage of incident database to design and operational software similar to Microsoft Word© spell check program. The database can run on the background of software and whenever certain keywords appear in the software interface, it is highlighted and the user can look into the incident information relevant to the keyword with the click of a button. The inspiration for this research came from Kletz's proposal and the second objective is dedicated to implementing this idea. Further motivation came from Bond's proposal and sheer explanation of the benefits.

Chung[2] has developed a standalone database based on fuzzy search and proposed integration of the database to computer tools used by plant designers, operators, and maintenance engineers, such as CAD and digital control systems. The two different types of data retrieval as referred to by Chung are manual and automatic. Plant designers are expected to use manual data retrieval at initial stage of design while automatic database systems will help them during the detailed design phase to identify and avoid past incidents. The automatic retrieval for designers is suggested by integrating database with flow sheeting or CAD systems. HAZOP teams are expected to make use of the manual system in order to identify potential hazards and operability problems that have not been effectively addressed in the design stage. While plant operators and maintenance personnel are expected to make use of only the automatic system, since they are less likely to browse through incident databases in their day to day jobs. For the benefit of the operators, the database could be linked to a digital control system, so that they could be further informed about a hazardous situation before an alarm goes off, which will give them the opportunity to react appropriately instead of just cancelling the alarm. For maintenance personnel, integrating database to a computerized permit-to-work system would provide them with the opportunity to identify hazards with the maintenance task at hand.

3. RESEARCH METHODOLOGY

3.1 Taxonomy Development

The incident taxonomy development method has been divided into three sections, as highlighted in Figure 3. Literature review of other incident classification and text mining of past incident information have been carried out to build an initial taxonomy. The taxonomy is then sent to five individuals, experts in the field of chemical process safety in order to identify inconsistencies in the taxonomy. The suggestions gathered from the experts were analyzed and changes made accordingly to develop the final taxonomy.

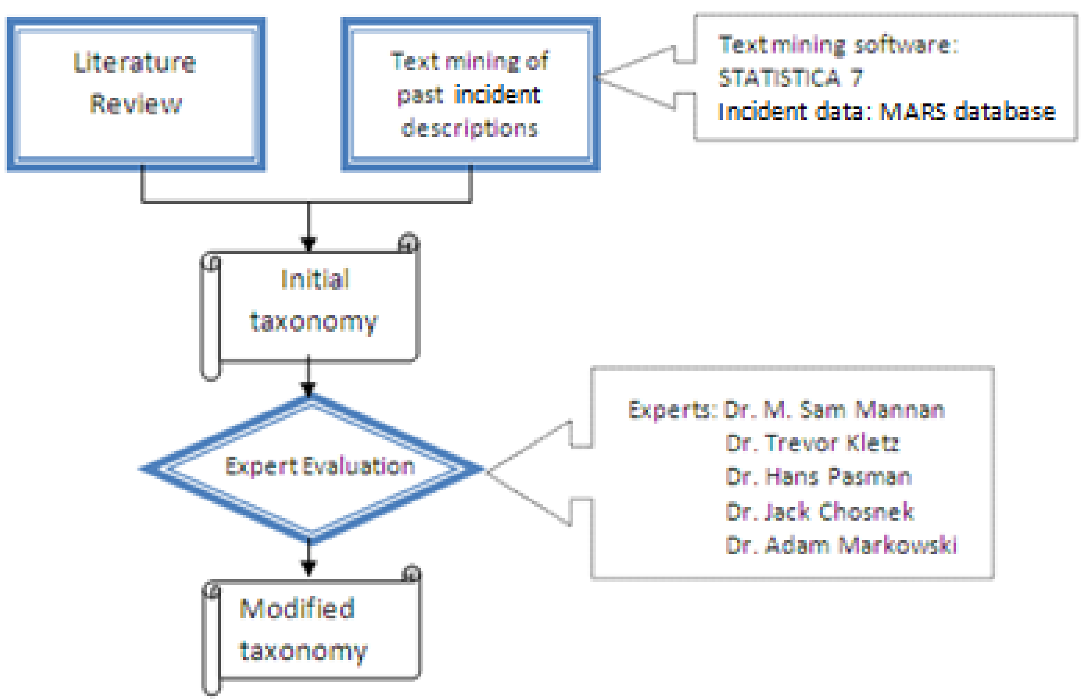

Figure 3. Applied methodology for incident taxonomy development

3.2 Database Search System

Once prepared, the taxonomy has been implemented as Microsoft Excel worksheets, where each node of the taxonomy tree is a folder. The taxonomy is then populated with the incident information from the MARS database short reports, classifying each incident in different folders according to their relevancy. This method is applied to create a database system, which has the option of guided search for the users by means of folder searching. The computer coding done to implement the taxonomy to a database system and to build an interface for the users to retrieve incident was developed using JAVA.

Besides folder search, the users are provided with the option of word search to retrieve incident information. Folder searching along with word searching will help narrow down the users' scope of search and direct them to more relevant results of their queries. A third approach is to render the database active through integration with a Management of Change software prototype.

The techniques applied for text mining past incident data is explained in detail in Sections 4.1 and 4.2. Section 5.1 provides details of the sources used to build the taxonomy. The method of building the search system is explained in Section 6.1, 6.2, and 6.3. Finally, methodology applied for integration of the MOC software prototype with the incident database is described in detail in Section 7.

4. TEXT MINING OF PAST INCIDENT DATA

Text mining is the process of deriving meaningful patterns, trends, and hypothesis from a collection of texts, that are not explicitly revealed in the text sources.[20] Text mining of past incident information has been performed for the purpose of finding unique words that represent the incident descriptions and also to reveal relationships between different keywords. Text mining is applied in the hope of finding related keywords that would aid in the taxonomy development process. The flow diagram for the text mining method is shown in Figure 4 and it is explained further in the rest of this section.

4.1 Data Preparation

Past data chosen for text mining is the MARS database, which have short incident reports. This database is particularly chosen for its numerous text fields, which can help in understanding not only trends but details of causes and lessons learned for each incident. A JAVA program has been coded to capture the MARS short reports available online from their website[10] and convert the HTML data into an Excel file format. A list of the variables captured is shown in Table 1. Some of the categorical variables in the MARS database have subcategories, which have not been absorbed in the dataset to avoid coding a complex JAVA program that 'crawls' or scans the MARS database to retrieve information. For example, the variables *Release*, *Fire*, and *Explosion* have subcategories of major occurrences (*solid release to air*, *flash fire*, *BLEVE*, etc.) and initiating events that fall under these categories. All of these variables have been listed as binary categorical variables for simplicity. The categorical variables describing the type of hazardous substances involved (*toxic*, *very toxic*, extremely flammable, oxidizing, etc.) have also been simplified from the original data. MARS database has a list of the type of substances classified according to the Dangerous Substances Directive and other European Union laws concerning chemical safety.[21] To keep it simple, categories such as *toxic substance* and *very toxic substance* has been classified as one and represents the binary variable *Toxic*. Other simplifications also involves avoiding subcategories of variables, such as *remedial measure* (includes subcategories *decontamination*, *restoration*, *other*) and *lessons learned theme*, which have been recorded as binary variables.

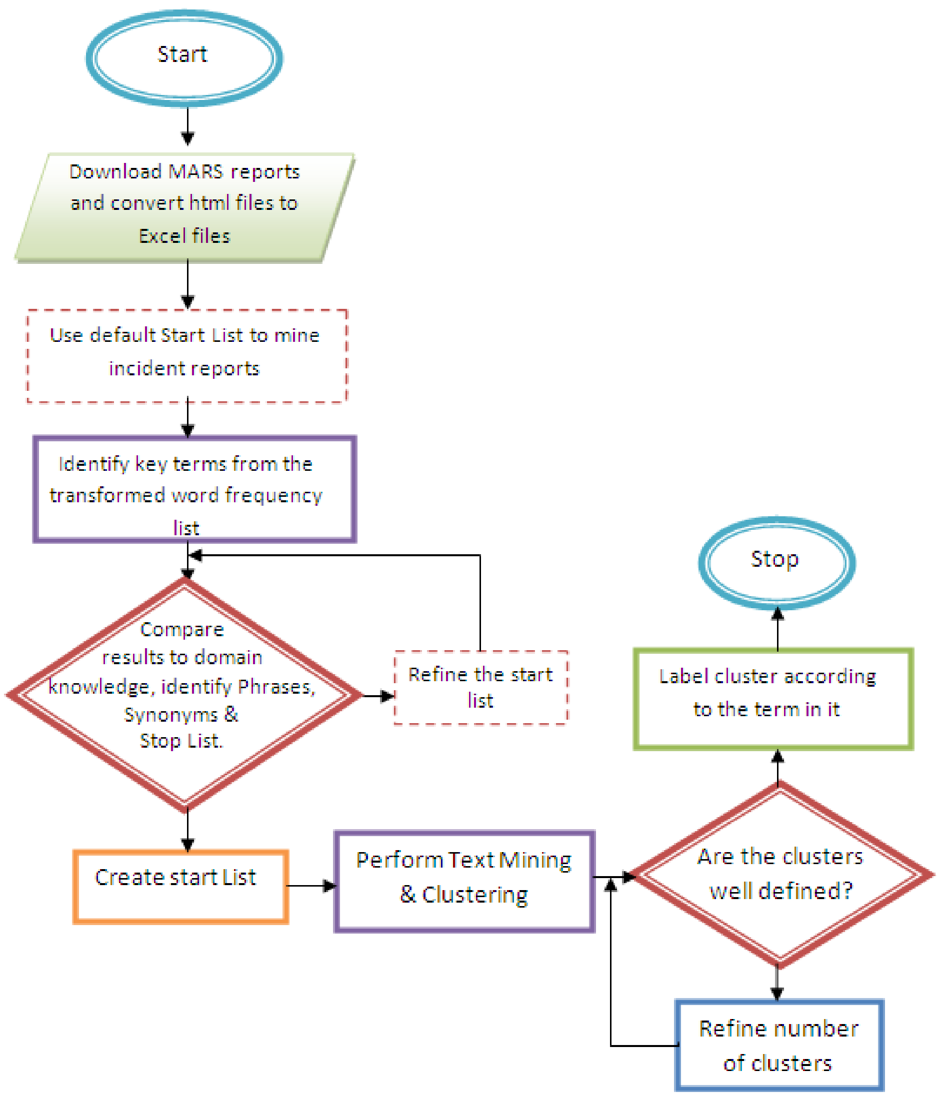

Figure 4. Flow diagram showing the Text Mining process of past incident data (Adapted from Raja & Tretter)[22]

Table 1. List of Variables from MARS Short Reports

Variable Name	Type	Description
Start Date	Text	Starting date of incident
Start Time	Text	Starting time of incident
Finish Date	Text	Finishing date of incident
Finish Time	Text	Finishing time of incident
Accident Title	**Text**	**Title of the accident description**
Accident Type	Categorical	Major accident, near miss, or other event as defined
Reported under	Categorical	Reported under Seveso I or II directive or other legislative requirement
Seveso II status	Categorical	What status of Seveso II directive it falls under
Industrial Activity	Categorical	Type of industry/industrial activity
Accident Description	**Text**	**A brief description of the incident**
Release	Categorical (binary)	1 if the incident involved release of a substance; 0 otherwise
Fire	Categorical (binary)	1 if the incident involved fire; 0 otherwise
Explosion	Categorical (binary)	1 if there was an explosion; 0 otherwise
Site Description	Text	Description of the site
Installation/Unit Description	**Text**	**Description of the installation, unit, or equipment involved**
Storage	Categorical (binary)	1 if the unit type was storage; 0 otherwise
Transfer	Categorical (binary)	1 if type of unit or operation is tranfer; 0 otherwise
Transport	Categorical (binary)	1 if type of unit or operation is transport; 0 otherwise
Process	Categorical (binary)	1 if type of unit or operation is process; 0 otherwise
Other Equipment	Categorical (binary)	1 if unit/operation does not fall on last four categories; 0 otherwise
Substances Involved	**Text**	**List and a brief description of hazardous substances involved**
Toxic	Categorical (binary)	1 if toxic substance is involved; 0 otherwise
Flammable	Categorical (binary)	1 if flammable substance is involved; 0 otherwise
Explosive	Categorical (binary)	1 if explosive substance is involved; 0 otherwise
Oxidizing	Categorical (binary)	1 if oxidizing substance is involved; 0 otherwise
Dangerous for Environment	Categorical (binary)	1 if the substance is ecotoxic; 0 otherwise
Causes of the accident	**Text**	**A brief description of causes of the incident**
Organizational	Categorical (binary)	1 if cause of incident is related to organizational issues; 0 otherwise
Plant/Equipment	Categorical (binary)	1 if cause of incident is due to equipment or technical failure; 0 otherwise
Human	Categorical (binary)	1 if cause is human error; 0 otherwise
External	Categorical (binary)	1 if cause is due to external events (eg. bad weather); 0 otherwise
Other Cause	Categorical (binary)	1 if cause does not fall on any of the last four categories; 0 otherwise
Consequences	**Text**	**A brief description of the consequences**
Human Consequences	Categorical (binary)	1 if fatalities, injuries, or other human consequences; 0 otherwise
Disruption	Categorical (binary)	1 if disruption of public activities, evacuation etc.; 0 otherwise
Environmental	Categorical (binary)	1 if incident resulted in pollution and environmental damage; 0 otherwise
Emergency Response	**Text**	**A brief description of the emergency response**
Remedial measure	Categorical (binary)	1 if remedial measure has been taken; 0 otherwise
Lessons learned theme	Categorical (binary)	1 if theme of lessons learned is identified; 0 otherwise
Lessons Learned	**Text**	**A brief description of the lessons learned**

4.2 Text Mining Method

Statsoft® Statistica software is used to perform the text mining applications. Statistica text mining algorithm process unstructured text data and convert unique words to meaningful numeric indices in order to make the information contained in the text available to different data mining algorithm.[23] Before the numeric index can be developed to give meaningful results, several pre-processing steps need to be done in order to prepare the data.

4.2.1 Pre-processing Steps before Indexing

Various pre-processing steps performed before the text is converted to numbers are described below:

Stop words

Before the index is built, Statistica text miner allows the option to choose *stop words*, which are words that would be excluded from the index. Statistica has an in-built English stop words list, which includes words such as prepositions that tend to exist in all documents and do not aid in characterizing individual document. Besides the in-built stop words, there is the option to add more unimportant words to the list.

Start words

Just opposite to stop words, *start words* are used to make sure the important words are included for indexing.

Phrases & Synonyms

Besides stop and start words, one can input a specific *phrase* to be indexed as a single term. For example, the term ‘source’ can mean source of a release and at the same time it can exist in the term ‘ignition source’, which has a completely different meaning. Therefore adding the term ignition source in the phrase list allows the term to be considered separately for indexing. A list of declared *synonyms* counts two or more words with similar meaning only once for indexing. Sometimes when it is difficult to work with a long list of words, synonyms can be effectively used to reduce this list by declaring words with similar attributes as one.

Stemming

Stemming is another important pre-processing step before indexing. Stemming reduces the words to their roots so that different grammatical forms or verb conjugations are identified and indexed as the same word. For example, call, calling, and called would all be considered one word after stemming. The stemming algorithm is language dependent; therefore the English stemming language has been used for this research.

Besides these preprocessing steps, there are other features in Statistica text miner that help to refine the selection of words from the collection of documents (corpus). These are:

- Excluding certain characters, numbers, or character sequences
- Excluding words that are shorter or longer than a certain number of letters
- Excluding rare words that occur in a small percentage of documents.

4.2.2 Indexing

Statistica indexer forms a matrix, which has the dimension of (total number of unique words) x (total number of documents). The words are selected after the preprocessing steps are performed and indexed according to their frequencies in a given document. This measure is called word frequency (*wf*) or term frequency (*tf*) for those terms that have been declared in the phrase list during preprocessing steps. Term frequency is the simplest and not the most effective measure to identify unique words. Thus the raw word frequency can be transformed to other measures.

Transforming word frequencies

Once the words are indexed, there are a number of ways to transform the raw word frequencies in order to summarize the extracted information. For this research the chosen transformation is inverse document frequency (*idf*). $idf_{i,j}$ for word *i* and document *j* is defined by the following equation:[20]

$$idf(i,j)=\begin{cases}0 & if\ \ wf_{i,j}=0\\ (1+\log(wf_{i,j}))\log\dfrac{N}{df_i} & if\ \ wf_{i,j}\geq 1\end{cases} \tag{1}$$

Where N = no. of documents

df_i = no. of documents where word i exist

$wf_{i,j}$ = frequency of word i in document j

df_i and $wf_{i,j}$ are also called document frequency and word frequency respective. The log function dampens the simple word frequencies and also includes a weighting factor that becomes zero if the word appears in all documents. Thus, this transformation is done to create indices that both reflect relative frequencies of words as well as their semantic significance over the documents.[23]

4.2.3 Latent Semantic Indexing by Singular Value Decomposition (SVD)

Singular value decomposition is an analytic tool for interpreting the meaning described by the extracted words through mapping the words or documents into a common space. Latent semantic indexing allows identification of underlying dimensions into which the words and documents can be mapped. That way the underlying or latent themes described in the incident reports can be identified.

The purpose of SVD is to reduce the overall dimensionality of the (words) x (documents) matrix to a lower-dimensional space. Each of these dimensions represents the largest degree of variability between words or documents. Ideally two or three most significant dimensions can be chosen that account for most of the variability. Statistica SVD is performed by the described process:

If $\boldsymbol{A}$ represents an $m\ x\ n$ matrix, where m is the number of documents and n is the number of words, SVD computes the $m\ x\ r$ orthogonal matrix $\boldsymbol{U}$, $n\ x\ r$ orthogonal matrix

V, and *r x r* matrix ***D***, so that ***A*** **=*UDV′***, and so that *r* is the number of eigenvalues of ***A′A***.

An iterative method is used to compute SVD in order to reduce dimensionality of the large and sparse matrix ***A***. This method produces accurate values for the large singular values but less so for the smaller singular values. Hence a 'scree' plot is produced to display relative sizes of the singular values and the values that are left of the elbow of the scree plot represents most of the variability and are retained for analysis. Scree plots for the MARS data can be found in Section 3.3. The word coefficients calculated by the program are computed as the matrix ***W***, where ***AW* = *U*** and matrix ***U*** is the document score matrix.[23]

One application of SVD is to find the word importance, which shows the relative sizes of the square root of the diagonal values in ***VDU′UDV′* = *VDDV′***. These indices can be interpreted as the extent to which individual words are represented by singular values and hence how important the words are for defining the latent space.[23]

4.2.4 Selecting Predictor Words with Feature Selection and Variable Screening

Feature selection and Variable Screening module (FSL module) in Statistica text miner allows selecting predictor words from a large list of indexed words. This is ideal for classification problems, where selection of a subset of predictors from a large list of candidates is necessary. Therefore, this tool can serve to select manageable number of predictors that are likely to be related to the dependent variables of interest.

The words converted to numeric indexes can be represented as continuous variables, where the variable is dependent on each document in the (term) x (document) matrix. Using FSL module then the predictor terms can be narrowed down for a given term. The FSL module computes the ratio of the variance between a given term (dependent variable) and predictor variable to variance within the dependent variable. It divides the range of values in each predictor into k intervals, where k is 10 by default to adjust the algorithm sensitivity to linear or non-linear relationships.[23] The user defined number of best predictors can be sorted using either the F or p value as criterion of predictor importance. For categorical dependent variables, the criteria for selecting best predictors are Chi-square and p values.

4.2.5 k-means Clustering

Once the Singular Value Decomposition is performed to reduce dimensionality of the input matrix, the top components can be used to analyze the data using different data mining techniques. One such technique is cluster analysis, whereby the incidents can be classified into similar clusters, based on word counts and relationship between words.

k-means clustering is a clustering technique where given a fixed number of clusters, the algorithm assigns observations to those clusters, whose means are as different from each other as possible.[20] The difference between the observations is evaluated with any one of several distance measures. For this research the simple Euclidean distance measure has been chosen.

4.3 Text Mining Results and Discussion

The text variables *accident title*, *accident description*, *installation/unit description*, *substances involved*, *causes of the accident*, *consequences*, *emergency response*, and *lessons learned* have been used for text mining. A list of stop words has been developed to eliminate all the unwanted words. A list of important phrases has been developed to include the common terms for this corpus of documents, such as 'boiling liquid expanding vapor explosion' or 'toxic cloud'. Also a list of synonyms has been built, which includes not only synonyms but also words with similar connotations, such as *people, population,* and *residents* (which most likely in the report, relates to the casualties or vulnerable population outside the fence) were all identified as one word *people*. All three lists were changed iteratively until a smaller group of words were chosen for indexing. Out of this group a start word list is developed by hand picking those words that contain more meaning in relation to the context of incident reports. These four lists, as shown in Appendix A, were finally incorporated to obtain a representative group of words for indexing. The total number of words indexed was 332 (reduced from around 1000 words) from a total of 696 incident reports.

The word frequencies for the indexed words were converted to inverse document frequency before running the SVD module of Statistica. The number of components obtained was 22 as shown in Figure 5. The top four components left of the elbow of the scree plot, are the salient ones that account for 31% of the variability between the words and the documents.

Figures 6 and 7 show the plots of component 1 against component 2 and component 1 against component 3 respectively, representing words in semantic space. The outliers in these plots are the significant or important words in the corpus of incident reports and the words appearing closer to one another are related to one another.[23]

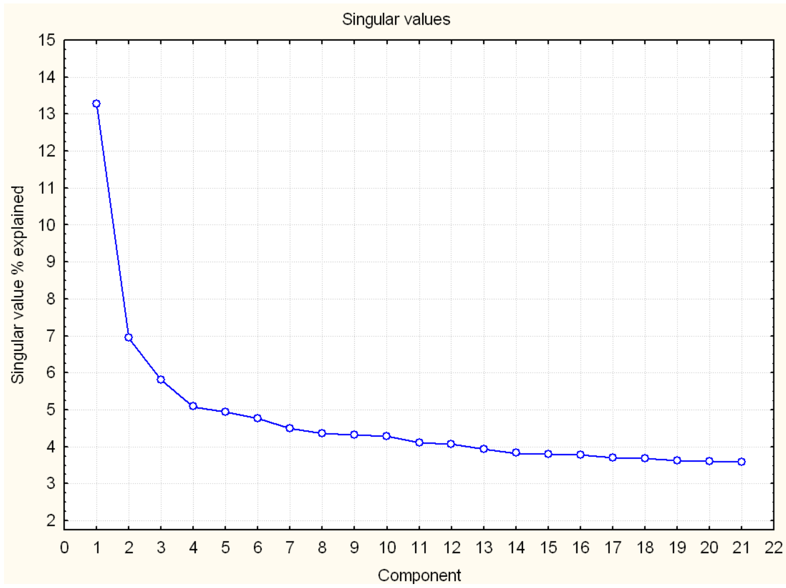

Figure 5. Scree plot

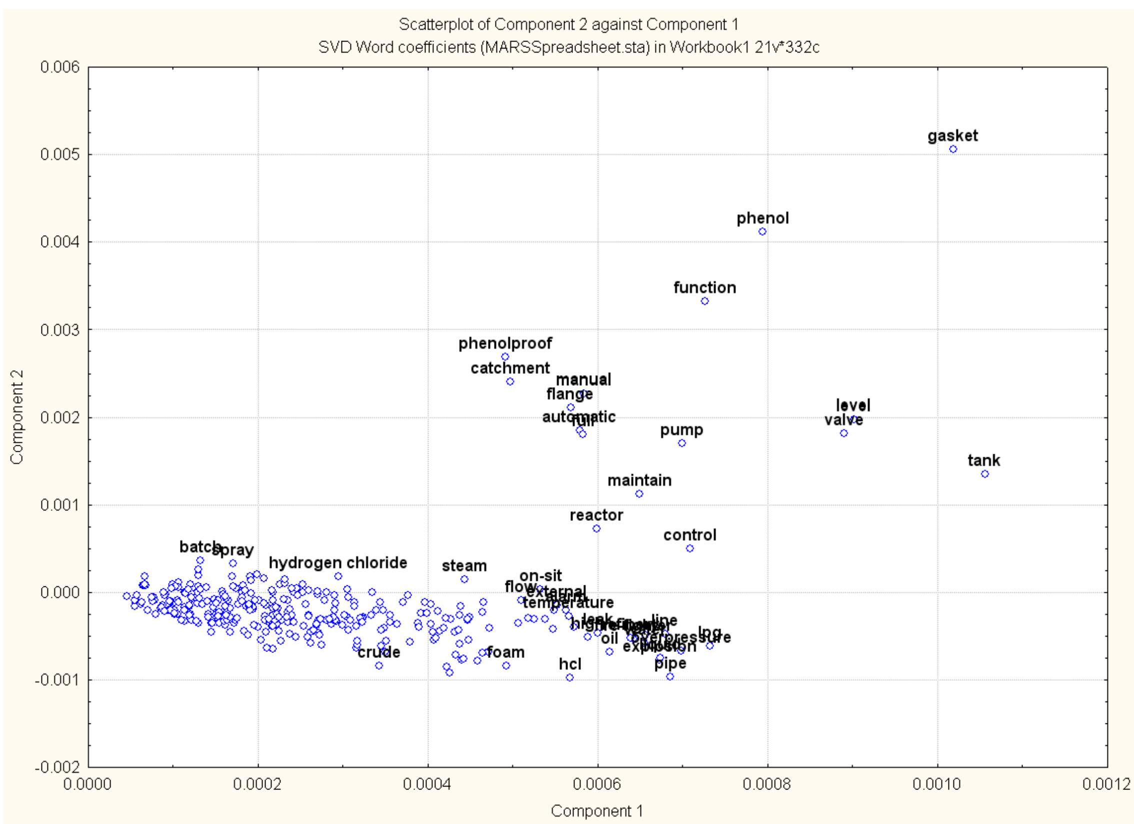

Figure 6. Words in semantic space: SVD component 1 vs. component 2

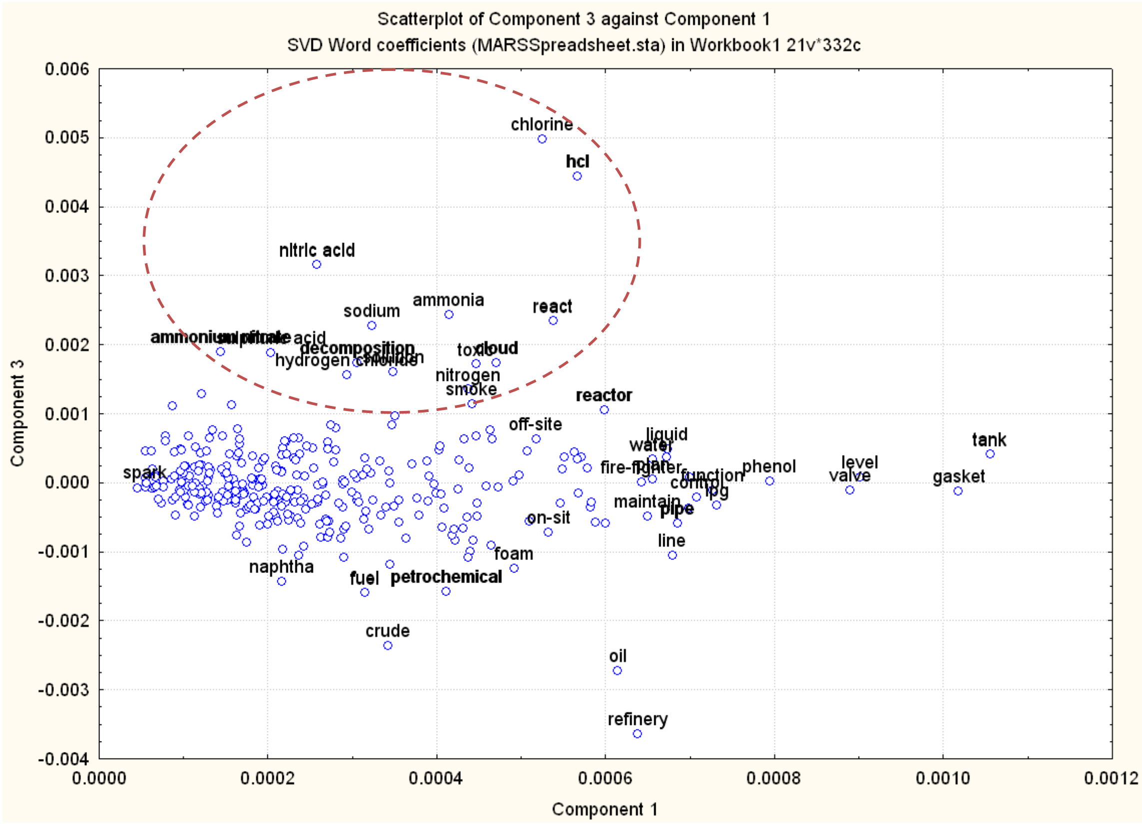

Figure 7. Words in semantic space: SVD component 1 vs. component 3

From Figure 6 it can be seen that the words *tank*, *valve*, *level*, *gasket*, *phenol*, *function*, *maintain*, *control*, *pump*, *automatic*, *manual*, *flange*, *catchment*, *phenolproof*, and *fill* are the outliers, which may indicate some of the reports include incidents where due to automatic or manual valve failure or pump failure the tank overfilled. It can also indicate events such as, due to inadequate maintenance, the pump flange or control failed leading to overfilling. Reviewing the incident database shows that there exist specific incident of phenol leak from pump due to inappropriate flange installation, that is not phenol-proof and thus reacted with the phenol. The phenol was collected in a catchment pit. The term maintenance has been listed as synonyms of maintain and maintaining, thus the word maintain suggest maintenance actions.

Figure 7 manifests existence of incidents related to refinery and petrochemicals. The group of words (circled in Figure 7) on the upper left corner indicates the ingredients of a toxic cloud. Fire in a fertilizer storage area can initiate decomposition of ammonium nitrate and result in release of toxic substances such as nitrogen oxide, ammonia, and nitric acid to form toxic vapor cloud.

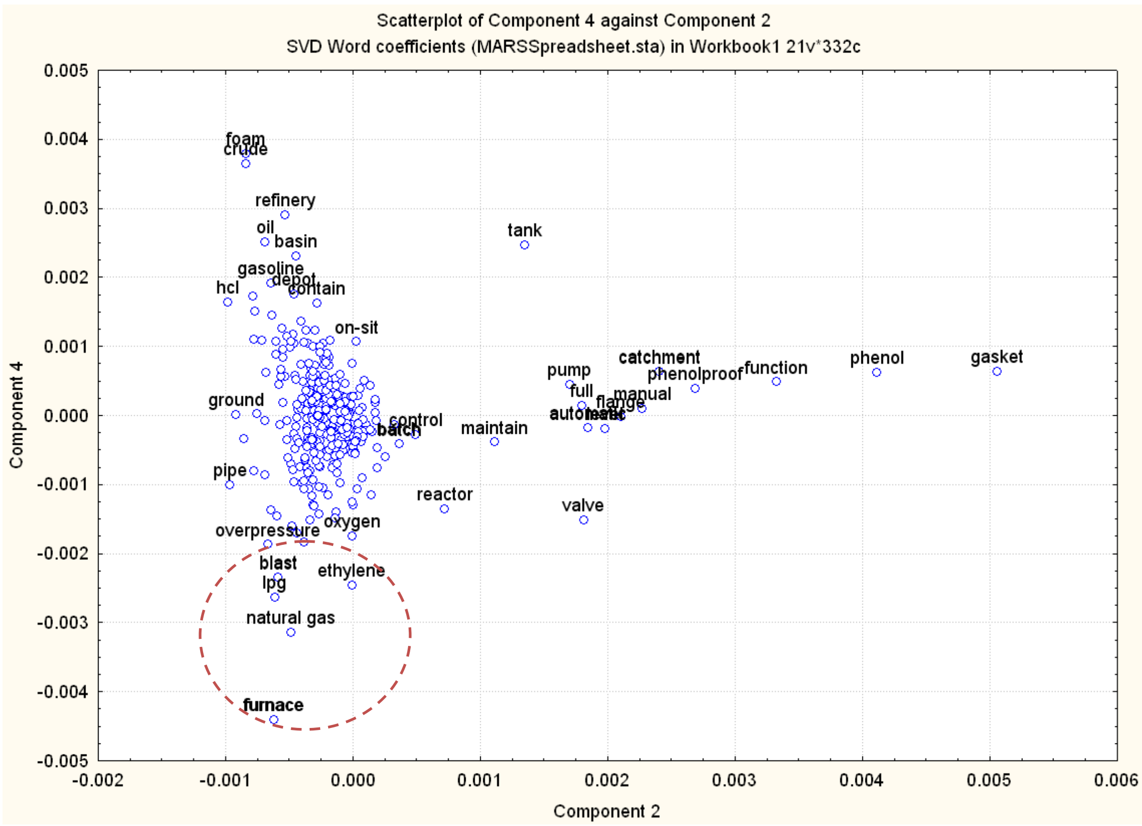

Figure 8. Words in semantic space: SVD component 2 vs. component 4

Another significant set of words is found in the plot of component 2 against component 4, as shown in Figure 8. These words may indicate incidents related to blast furnace, gas container failure and resulting fire, or gas container failure resulting in nearby natural gas installation destruction and fire etc.

The words of interest were further drilled down using the 'feature selection' tool in Statistica, as shown in Figures 9-15. The importance plots were obtained for different words and also some binary variables. The significance of these plots is that the predictor words are likely to occur at the same document, where the dependent variables (word or binary variable) occur. Figures 9 and 10 show the importance plots for gasket and tank, the significant words that we saw common in the semantic space plots (Figures 6, 7, and 8). From these plots, the basic ideas that can be deduced is that tank failure can result from pump or valve failure with resulting event of overfilling and release. Bund or catchment pits can act as passive mitigation system to control or hold the released material. Automatic or manual valve failure and pump failure causes can be related to inadequate maintenance with specific reasons such as phenol reacting with gasket and causing leakage. Tank incidents can also be related to empty tanks or fast emptying of tanks causing a vacuum.

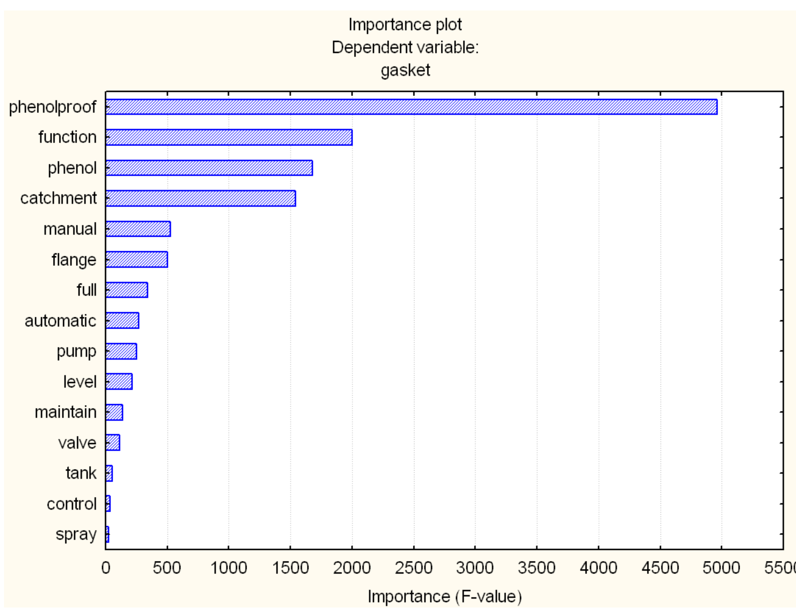

Figure 9. Importance plot for the word 'gasket'

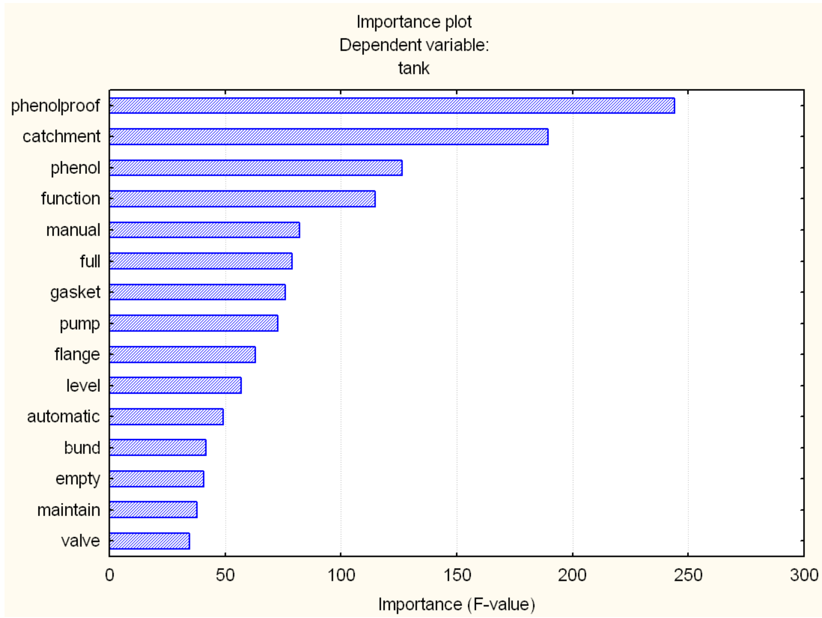

Figure 10. Importance plot for the word 'tank'

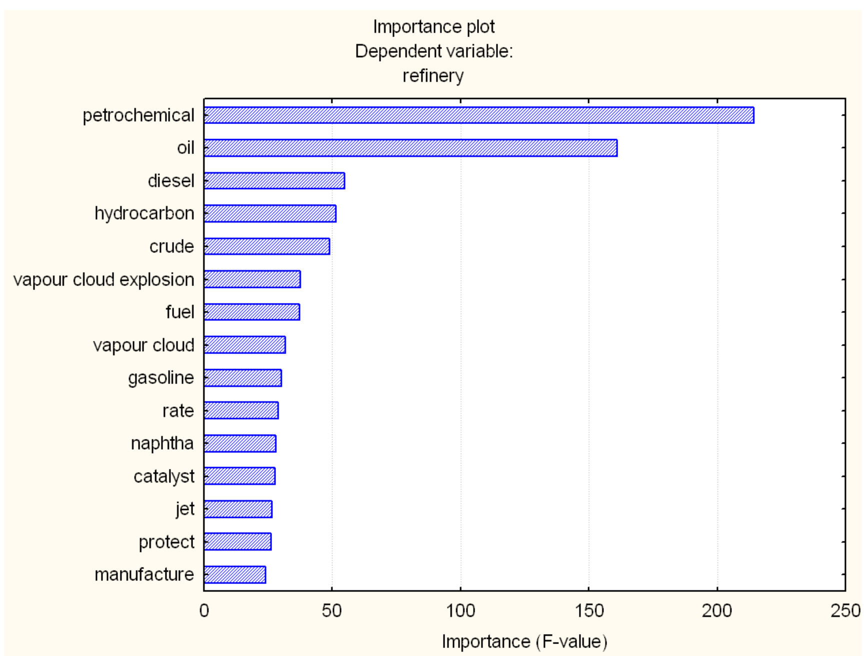

Figure 11. Importance plot for the word 'refinery'

Figure 11 shows the top 15 descriptive words for incidents in ‘refinery’ (word chosen from the outliers of Figure 8).

Hazardous materials, conditions, consequences, and mitigation measures of a possible toxic cloud formation incident can be identified from Figure 12.

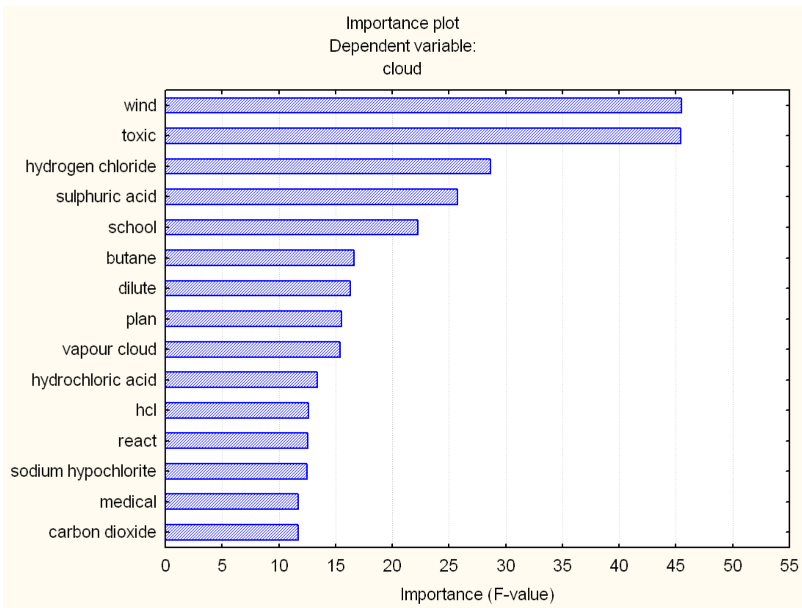

Figure 12. Importance plot for the word ‘cloud’

Figure 13 shows the keywords for possible conditions leading to overpressure and following consequences.

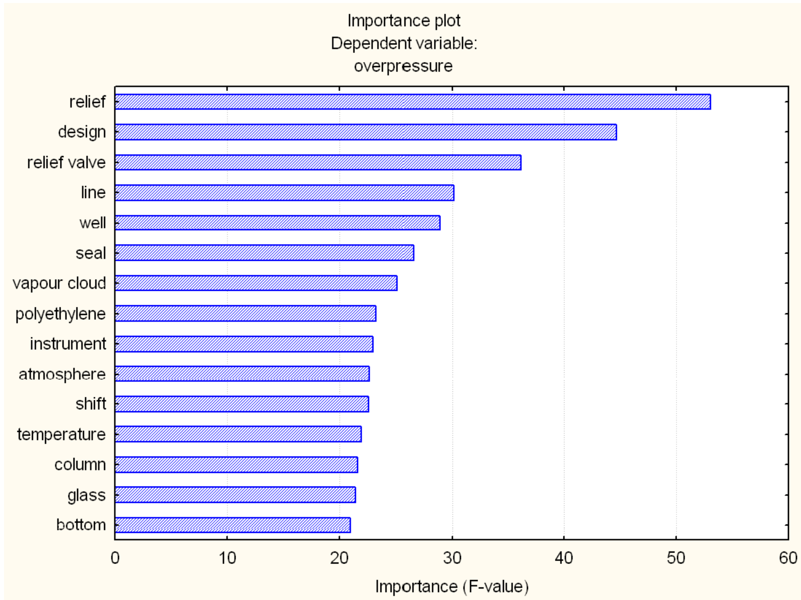

Figure 13. Importance plot for the word 'overpressure'

Various importance plots similar to Figures 9-13 have been obtained to analyze the information contained in the MARS data in order to find some informative relations for taxonomy building purposes. It should be noted that importance plot with binary variable as the dependent variable, sometimes provided a different result compared to that for a word variable. For example, the importance plots for the word 'release' and the binary variable 'release', as shown in Figures 14 and 15, are a little different. Release events usually result in a cloud formation, which may later result in an explosion; noticeably, both the terms, *explosion* and *cloud* are present in Figure 14. However, none of these words exist in the top 15 important descriptor words for the word 'release', although the term, *atmosphere* may imply flammable atmosphere. This shows that in a release event, which resulted in cloud formation or explosion, the word release is likely not used. This is one of the limitations of using only text variables for analysis.

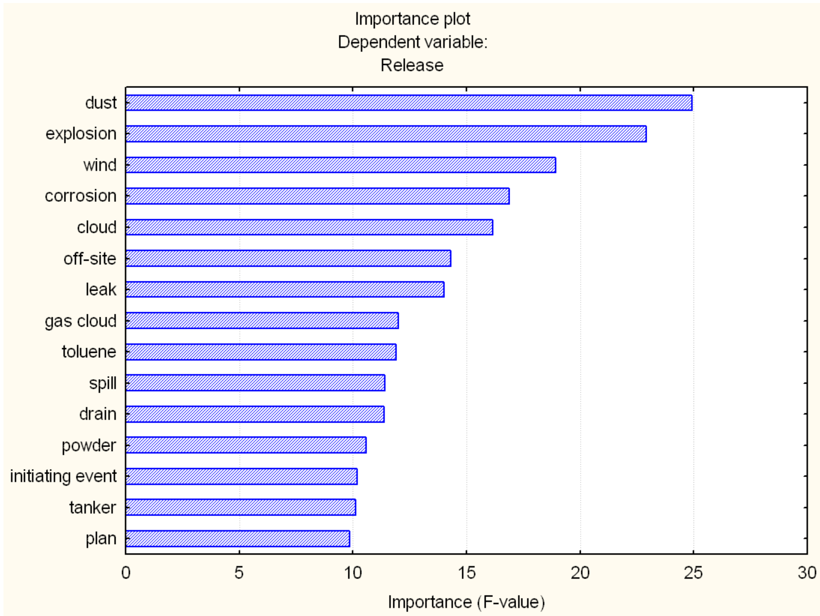

Figure 14. Importance plot for the binary categorical variable 'Release'

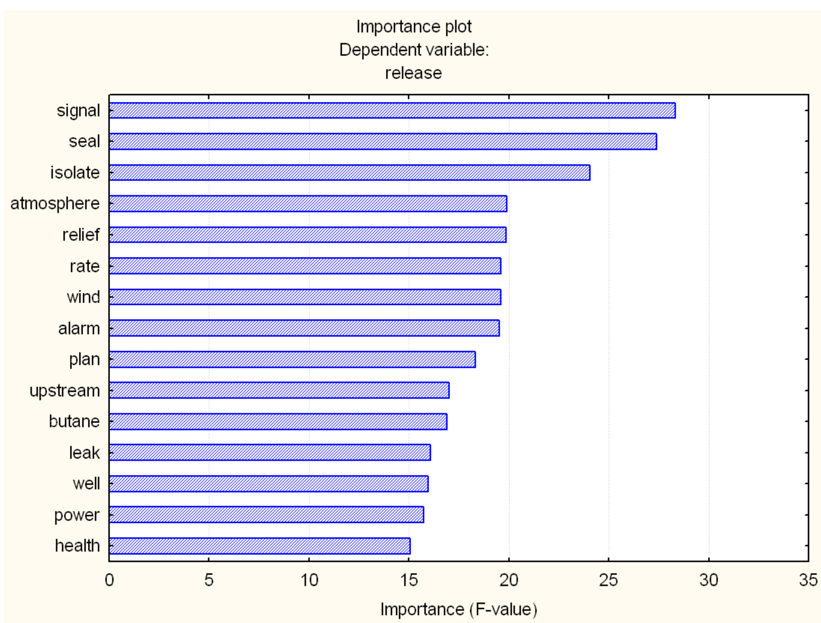

Figure 15. Importance plot for the word 'release'

Next, k-means clustering application has been applied to the SVDs for further analyzing the words and finding groups of words that are related and that make natural clusters.

Initially the total number of clusters chosen was 5 and went up to 12 clusters until some meaningful word clusters were obtained. The optimum number was 7 clusters.

Table 2. Clusters of All Text Variables with Representative Terms

Cluster	Percent	Representative Terms
1	45%	boiler, oxygen, solvent, ethylene oxide, methanol, catalyst, flammable liquid, hydrocarbon, burst, carbon dioxide, shock wave, sprinkler, runaway reaction, toluene, fire water, relief valve, flash fire, water curtain
2	8%	explosion, compressor, pipeline, powder, mix, air, roof, tower, shell, empty, dust, chamber, atmosphere, ignite, benzene, flame, electric, repair, wall
3	10%	reactor, gasket, tank, pipe, vessel, lpg, phenol, valve, line, leak, liquid, level, overpressure, temperature, high, phenolproof, fire-fighter, catchment, pump, full, water, release, alarm, maintain, control
4	20%	waste, bund, crude, drain, smoke, tanker, basin, sewage, relief, vent, road, isolate, hospital, health, environment, shut, rain, river, spill, rate, shutdown, people, exposed, clean, overflow
5	< 1%	ammonia
6	14%	natural gas, refinery, foam, oil, naphtha, seal, corrosion, gasoline, column, methane, fuel, propane, flare, sulphur, gas cloud, vapour cloud explosion, butane, design, rupture, collapse, petrochemical, extinguish, diesel, heat
7	2%	chlorine, furnace, blast, decomposition, nitrogen, nitric acid, ammonium nitrate, iron

Table 2 provides a summary of the representative terms in each cluster and the percentage of each cluster segment among the total number of words. It should be noted that the representative terms are not a complete list of term members for individual cluster. A detailed table of results for each cluster with all member terms can be found in Appendix B. Table 3 provides a summary of the cluster names or labels that are assigned to the clusters after examining the significant keywords in each distinct cluster. Since this analysis includes data from all applicable text variables in the MARS database, it is presumed that the cluster classification would incline more toward broad categories of incident types.

Table 3. Cluster Labels of All Text Variables

Cluster	Percent	Cluster label
1	45%	Flammable gas fire scenario and some toxic release
2	8%	Explosion (chemical & physical)
3	10%	Flammable liquid release with resulting fire
4	20%	Environmental/off-site consequences due to toxic release
5	< 1%	Ammonia incidents
6	14%	Flammable gas/vapor explosion
7	2%	Decomposition and runaway reaction

Cluster 1 includes mostly terms that relate to flammable gas fire, although there are some terms, such as *sulphuric acid*, *waste water*, *hydrogen fluoride*, *hydrogen chloride*, *sodium hypochlorite*, and *hydrochloric acid* that are toxic substances, which may indicate these substances were released due to fire as domino effect. Or these terms may simply relate to toxic substance release. Similarly, other clusters also contain a few outliers that do not exactly belong to the rest of the cluster members and thus, the classes defined in Table 3. This implies that the clusters are not clearly differentiated from one another. This limitation is partly due to the large number of words, which can be further reduced by declaring words with similar connotations as one; it is also perhaps a result of using all text variables instead of focusing on a particular one at a time, such as *causes* or *consequences* etc. The former is a tedious, repetitive process that requires domain knowledge as well as a clear knowledge of the dataset. Since the iterative process had to be stopped at one point, the pre-processing stage was discontinued after obtaining 332 terms, which is small compared to 1000 terms but not small enough to provide distinct, manageable clusters.

In order to obtain more meaningful result, similar text mining analysis has been applied for only the *causes of the accident* text variable. Total number of unique words obtained was 131 and total number of SVDs was 27, among which first four are the most significant. Figure 16 shows the words in semantic space for this study. Here similar terms such as *tank*, *valve*, and *gasket* are the outliers as in Figure 6, with the addition of some causal terms such as *degrade*, *inspect*, *maintenance*, *control*, and *procedure*.

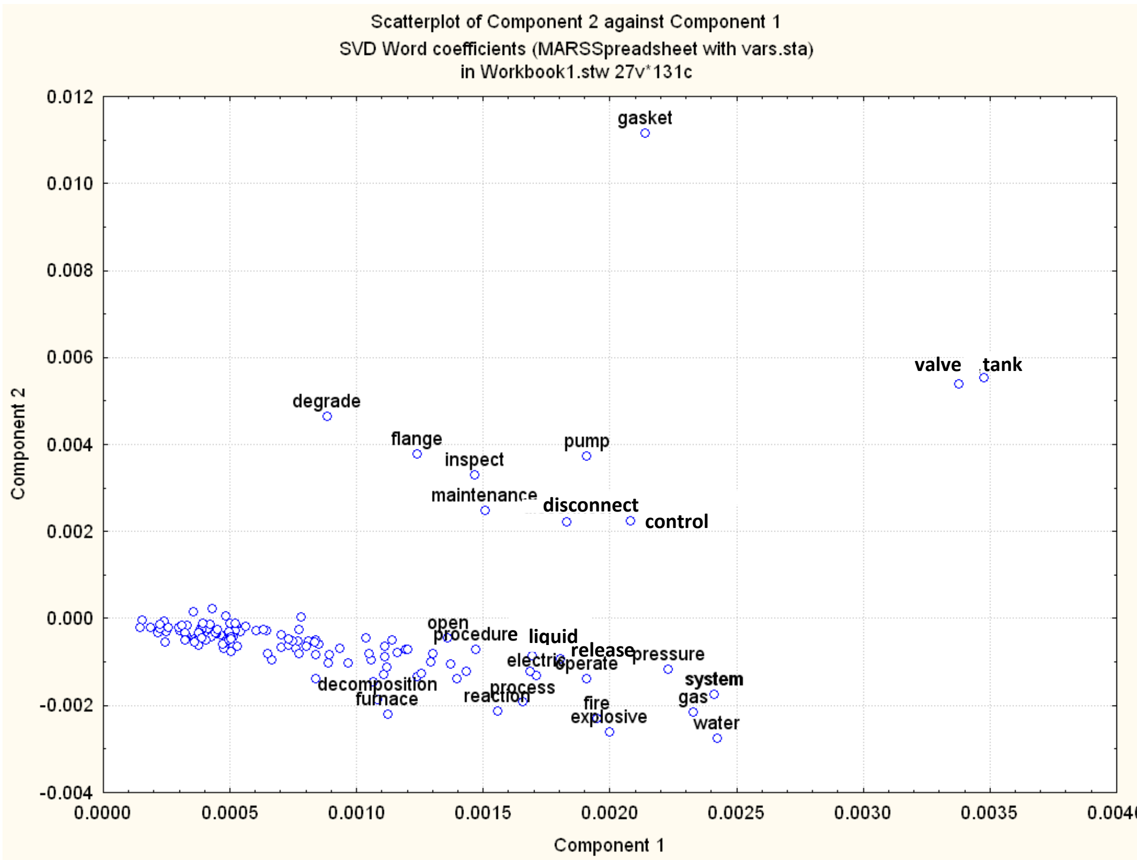

Figure 16. SVD Component 1 vs. Component 2 for Causes Variable

Clustering analysis has been further applied to the *causes* data and the results of seven clusters are shown in Tables 4 and 5.

Table 4. Clusters of 'Causes' Variable with Representative Terms

Cluster	Percent	Representative Terms
1	5%	air, electric, fire, open, safety, system
2	4%	alarm, high, pressure, supply, temperature
3	8%	analysis, corrosion, design, initiating event, inspect, internal, maintenance, material, metal, pipeline
4	6%	cloud, explosive, flammable, gas, mix, release, tank, vessel
5	8%	backflow, control, cool, gasket, hydrogen, process, reaction, valve, water
6	15%	clear, deficient, disconnect, drain, equipped, hazard, hose, instruct, leak, mechanic, operate, organizational, procedure, pump, repair, risk shift, train, work
7	55%	atmosphere, block, burst, change, chemical reaction, combusted, concentrated, decomposition, exothermic reaction, dust, expansion, external, friction, high pressure, high temperature, hot, impact, isolate, overfill, overflow, overheated, overpressure, oxygen, runaway reaction, toxic, motor, shaft, seal, rupture, uncontrolled, tube, instrument, hydraulic

Table 5. Causes Cluster Description

Cluster	Percent	Cluster description
1	5%	Electrical equipment, overpressurization of air reservoir and safety system failure causing fire
2	4%	Alarm failure or lack of operator intervention on alarm resulting in temperature and pressure rise
3	8%	Inadequate maintenance, design, or inspection causing equipment failure
4	6%	Accidental mixing, tank/vessel failure causes release of flammable gas
5	8%	Control system failure and reactive hazard
6	15%	Organizational causes and human error
7	55%	Combination of process upset, chemical causes, and equipment failure

Cluster 7 is the largest and consists of terms that relate to a combination of broad categories, process upset, chemical causes, and equipment failure. Clusters 1, 2, and 4 relate to very specific causes. Thus, cluster analysis on a specific variable can narrow down the study scope further to look into the detail of different areas of incident descriptions.

Though text mining of past incidents can provide promising results, further modification needs to be done on the pre-processing steps to have a cleaner data in order to obtain significant results. In addition text mining is a tool, which can only be applied efficiently, having comprehensive domain knowledge to guide the study. Therefore at this point we need to divert attention toward taxonomy development with knowledge from literature and using text mining results, wherever applicable.

5. TAXONOMY DEVELOPMENT

There are many reasons and objectives for the development of incident databases and the associated incident taxonomy. Depending on what the ultimate goal is, or what purpose it would serve, one or many different approaches can be chosen. Keeping the goal in mind - render incident databases more accessible to the users and thus more user-friendly, a potential taxonomy has been proposed in this research.

5.1 Initial Taxonomy Preparation

Initially the taxonomy has been prepared from literature review of other taxonomies related to process safety. The collection of literature used is described in Section 2.2. The initial taxonomy has been divided into six parts namely, chemicals, equipment, operation, causes, consequences, and barriers in order to represent the bow-tie as shown in Figure 2 in Section 2.2. The idea behind this division was that besides helping users to find information of interest from any of these categories, it would also serve as a potential guide to scenario building. This may be helpful for users during Process Hazard Analysis (PHA) studies, where the first step is identification of potential hazards.

5.2 Expert Suggestions and Modifications

Some experts have provided detailed suggestions for change, while others have suggested minor changes, such as addition of another node on top of the existing taxonomy. It has been observed that different ideas are offered by different reviewers depending on individual interest and contemplated goals of this research. This observation supports the philosophy that there is no ‘the way’ of preparing incident taxonomy. Two of the major interests of reviewers and the related suggestions are described in Sections 5.2.1 and 5.2.2.

5.2.1 Taxonomy for Incident Scenario Building

For the incident taxonomy to be useful for scenario building, the *causes* and *consequences* taxonomy has to follow a sequence of events. For instance, the *causes* taxonomy should follow more like the fault trees and the *consequences* taxonomy like the event trees of the ARAMIS project, as described in Section 2.2. This is because knowing what undesirable event can lead to what critical event via what possible routes is essential to determine the recipe of a potential disaster. Likewise knowledge of what critical event can lead to what dangerous phenomenon via what routes is necessary to determine the consequences. Hence, having the taxonomy arranged this way can aid the user to build a full picture of the incident scenario (ref. personal communication with H.J. Pasman).

5.2.2 Incident Taxonomy for Incident Investigation

Incident taxonomy could also be used as a guide in incident investigation. During an incident investigation, the list of keywords in the taxonomy, arranged in a hierarchy, can serve as a checklist for the investigators, who need to know what they should be looking for in order to establish causes or identify causation. One of the biggest problem of information retrieval is people sometimes do not know exactly what they want and even if they do, they may not know how to ask for it or use the same descriptors to ask for it. Therefore having a taxonomy provides an educated guess for the investigators. The main benefit is however, people will add keywords to their incident descriptions, which may help other people to pick and choose what keywords to use and find exactly what information they need (ref. personal communication with J. Chosnek).

5.3 Taxonomy Result for Incident Database Search System

The initial taxonomy with some modifications from expert suggestions that led to a detailed taxonomy can be found in Appendix C. However, from aggregating expert comments, it has been observed that in order to have a taxonomy that serves the purpose of a more user friendly database, few important considerations have to be encompassed. These are:

- The person in charge of inputting information in the database will not spend significant amount of time choosing the right keywords, if given a laundry list of choices. Therefore the choices need to be limited.
- The more keywords that are available and used, the narrower the query result, however important information may be missing out from this result. On the other hand, if too few keywords are used, the number of hits is going to be too large and the user faces the challenge of fishing for relevant information from unrelated ones.
- The provider of information as well as the seeker needs to have the same understanding of the meaning of each keyword. One way is to define each keyword. Fortunately, the benefit of having a hierarchical taxonomy is that, the inner keywords can serve as examples of the outer ones (inner refers to those that are deeper down the taxonomy branches). And a very narrow definition will be counterproductive, since this will require the need for numerous keywords to cover a broad category. Also, it is necessary to use RAGAGEP (Regularly and Generally Accepted Good Engineering Practices) terms to avoid misconceptions.

With these considerations, the detailed taxonomy in Appendix C has been further modified to reduce the number of keywords in order to meet the targeted purpose. The final, modified taxonomy can be found in Appendix D. Description of the six divisions, literature sources used and modifications are provided in Sections 5.3.1-5.3.6.

5.3.1 Chemicals Taxonomy

Chemicals branch of the taxonomy would categorize incident descriptions according to the chemicals that are involved in the incident. This branch has been classified into the *state* of the chemicals, *type*, *hazard classification*, and *use*. The hierarchy has mainly been gathered from Occupational Safety and Health Administration's (OSHA) Globally Harmonized System of classification and labeling of chemicals (GHS)[13] and the Mary Kay O'Connor Process Safety Center's unpublished research work on development of a general incident reporting taxonomy based on HSEES, NRC, EPA-RMP, OSHA, NIOSH (The National Institute of Occupational Safety and Health), DOT NFIRS and IPIS database taxonomies.

Modifications made to the initial chemicals taxonomy include conversion of the detailed hazard classification keywords to simpler terms. For example toxic substance health effects such as 'Serious eye damage' or 'respiratory or skin effects' has been eliminated and only the broader terms, acute toxicity and chronic toxicity, have been included. The modified taxonomy is shown in Appendix D, where the number of column corresponds to the number of level of the hierarchy. Hence, the first column is first level; the second column is second, etc.

5.3.2 Equipment Taxonomy

Equipment branch correspond to the equipment involved in the incident. It has been based on the taxonomies proposed by Chung and Jefferson[11], Anand[12], Markowski[14], and MKOPSC unpublished research on incident reporting system taxonomy development. The initial modifications of this taxonomy was based on the concept of scenario building, thus the equipment were classified in a detailed hierarchy. However, modifications done later reduced the breadth of the taxonomy significantly and listed equipment that is commonly involved in incidents at the outer levels. As per the expert suggestion, a sub category of ‘facilities’ was also added, which would contain information about the type of facility or installation.

5.3.3 Operation Taxonomy

Operation part of the taxonomy corresponds to the type of operations involved in the incident. This has been put together based on Chung and Jefferson[11] and Markowski’s[14] taxonomy. Major change made on this taxonomy is eliminating some of the subcategories under ‘processing’, which defines the processing steps or phase when the incident happened, such as ‘heat transfer’, ‘separation’, etc. This is because it is redundant here, since most of these subcategories are already included in the equipment taxonomy.

5.3.4 Causes Taxonomy

Causes branch of the taxonomy represent the causes of the incidents. The initial taxonomy breaks down data into different categories up to level 3. From level 4 onward the causes have followed the general fault trees developed in the ARAMIS project.[17] Although the ARAMIS fault tree starts from critical event and leads backwards in the sequence of events, all the way to undesirable event, en route of necessary and sufficient causes, direct causes, and detailed direct causes, for simplicity the initial taxonomy included only the applicable causes and not all the details. Besides the ARAMIS project, the information to build this taxonomy has been taken from Chung and Jefferson[11] and Markowski[14].

The initial taxonomy was created partially based on the idea of scenario building as future application. Thus the modifications made on this taxonomy were to reduce the breadth of the taxonomy and keeping the format more categorical as opposed to sequential in order not to compromise the primary goal – improving user accessibility of incident database. One of the valuable suggestions from the reviewers of this taxonomy was to include both root causes and initiating events or immediate causes in the *causes* taxonomy. For this reason, a subcategory of initiating causes has been added on top of the detailed *causes* subcategories. The detailed *causes'* subcategories in this case serve as the root causes. The initiating event mostly follows the categorization of CCPS-PSID.

5.3.5 Consequences Taxonomy

Consequences branch correspond to the consequences of the incident. For simplicity the first level of this taxonomy has both consequences of dangerous phenomenon, such as fire, explosion, etc. as well as consequences on people, plant/property/equipment, etc. It has been put together from Chung and Jefferson[11], Markowski[14], and ARAMIS project[17]. The explosions subcategory has been classified according to explosion classification of Abbasi et al.[24] Again the modifications done to this taxonomy were to make it more structured in order to meet the goals of this research. The subcategory, ‘internal fire/explosion’ (fire or explosion in a confined space, such as tank or vessel) has been taken out from the initial taxonomy, since this subcategory is redundant, as this is already contained in the fire and explosion subcategory and may create confusion among the users.

5.3.6 Barriers/Layers of Protection (LOP) Taxonomy

Barriers branch correspond to the barriers that either eliminate or prevent a hazard from turning into a critical event, protect the facility by taking early action, and if the critical event has to happen, barriers help mitigate or reduce the consequences (as in barriers in bow-ties). The barrier taxonomy has been kept simple and categorized according to their type. Markowski’s[14] taxonomy has been followed to build this structure. In addition, Sklet’s[25] definition and classification of barriers has been studied to make a better informed decision of dividing the barriers into prevention, protection, and mitigation barriers. A fourth subcategory, 'avoidance' have been suggested by the reviewers, which may include the inherently *safe* designs (differentiated from inherently *safer* design) that have been able to avoid or eliminate the hazard completely. However, this change has not been made, since 'inherently *safe* design' can be incorporated into the 'prevention' subcategory in order to follow the target of keeping the keywords to a minimum.

Further changes done to the taxonomy includes modifying the broad category name 'barriers' to Layers of Protection, since it is a more preferable RAGAGEP term. Questions has been raised regarding the usage of the LOP taxonomy, since it is highly unlikely for incident reports to contain information on whether an existing LOP prevented the incident from becoming a worse one. However, the experts agree that it is a good idea to have LOPs as a keyword, which may force the community to think about 'what went well' besides 'what went wrong', when doing an incident investigation.

5.3.7 Other Modifications

Besides all the amendments to the initial taxonomy, a seventh part of the taxonomy has been added as per Dr. Trevor Kletz's comments. The seventh taxonomy is 'Actions and Recommendations' to include the recommended changes or actions taken after the incident. This can improve ways of operation and design and prevent recurrence of the incident. In order to reduce the breadth of the taxonomy, a maximum of three levels has been suggested with the option of further levels that can contain examples. However, the modified taxonomy does contain some information at the fourth level, although in rare occasions, to maintain a coherent hierarchy.

As a potential for future work of converting the database to an incident scenario building application, an example has been set by linking some nodes of the main taxonomy to that of the ARAMIS fault trees and event trees. This can lead the users to the other causes and the consequences related to a fault or event. For this research, only six such fault tree nodes have been linked to display potential of this linkage. The format of the fault/event tree is still under consideration. An image embedded in an html page is used at this moment but this may not be feasible for really long fault trees. The method used for linking this data is explained in Section 6.3.

6. SEARCH SYSTEM DEVELOPMENT

Some of the MARS incident database reports (html pages) have been manually classified in the modified taxonomy nodes to prepare for the folder search system. An information retrieval system was developed based on the populated taxonomy. To make it easier for the end-users to retrieve information using word search as well as folder search option for their query, a user interface (UI) was designed. A third option is word search within a folder search, which can greatly reduce or narrow down the search results for a given query and at the same time provide more relevant results. Description on the design and development of the search system is provided further in Sections 6.1-6.3.

6.1 Word Search

The JAVA code used for word search can be found in Appendix E. The following is a list of the tasks performed to prepare the word search module:

- All the MARS database pages were downloaded.
- The HTML pages were parsed to retrieve the text information that is displayed on the page. Parsing is the process of breaking down text into tokens or words that can be used for indexing. The resulting content was saved as text files.
- An index is generated using the Apache Lucene[26] parser on these text files.
- Lucene search functionalities are used on this index based on the user query to find the relevant results.

6.2 Folder Search

The JAVA code used for the folder search can be found in Appendix E. Here is a list of the tasks done to prepare the folder search module:

- The taxonomy was prepared as a Microsoft Excel sheet.
- The Excel Workbook was parsed and converted into a string data (sequence of characters) which can be recognized and converted to a folder tree by the UI module.
- The string is returned to the invoking function (UI module).

The folders at level 3 inside the *causes* branch are clickable that pops up images of ARAMIS fault trees, relevant to the branch. As mentioned in Section 5.3.7, for this research only a total of six images of ARAMIS fault tree are linked to the causes branch.

6.3 Word Search within a Folder

The JAVA code used for the word search within a folder can be found in Appendix E. The word search functionality for this part is as used in Section 6.1. The function takes two parameters, one is the name of the folder that is currently selected, and the other is the search query. The word search functionality can narrow down result for a long list of data, when used within folders.

The functioning of the search system is further depicted in Figure 17 and Appendix F contains screenshots of the running database application.

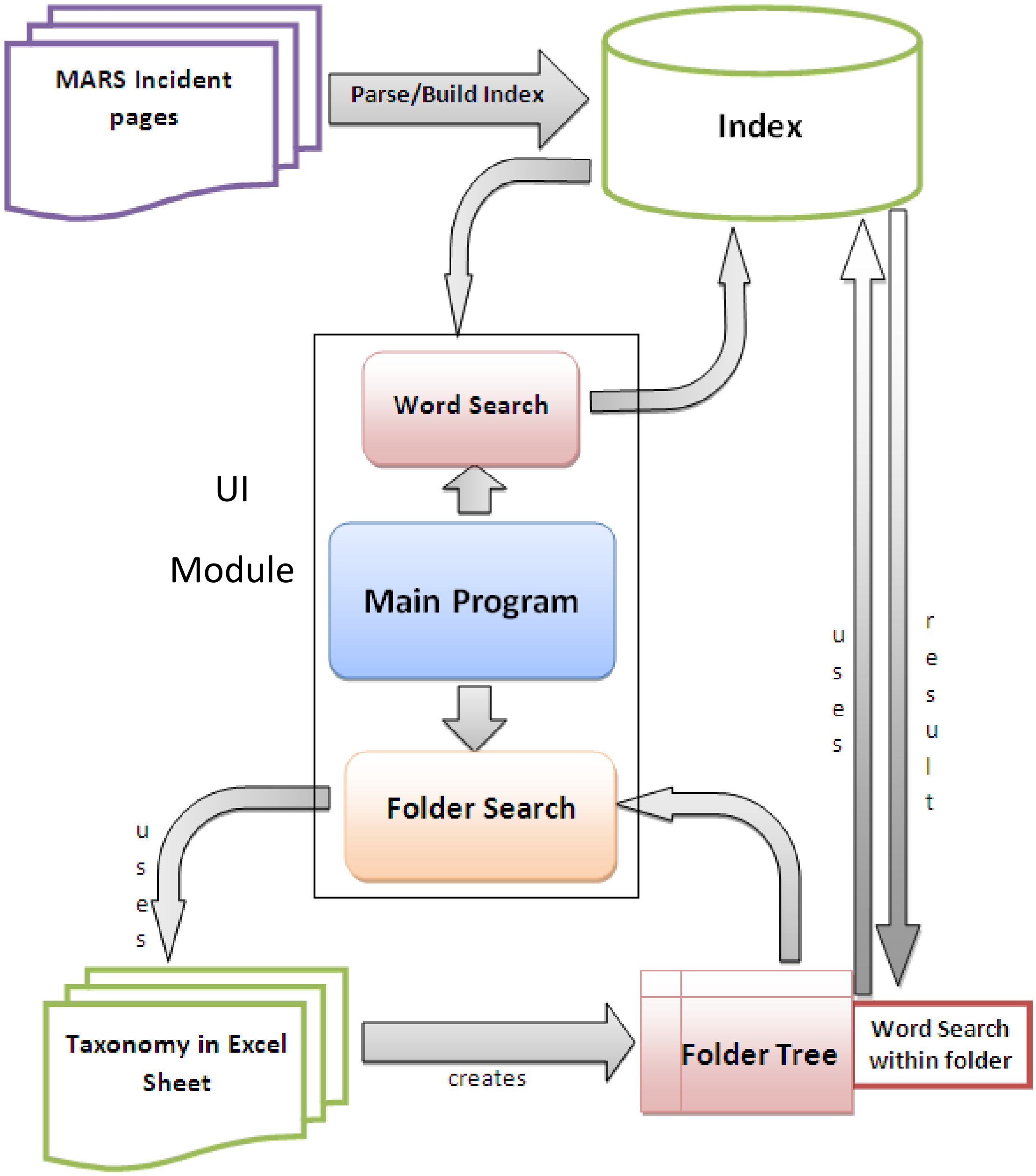

Figure 17. Flowchart depicting how the search system works

7. INCIDENT DATABASE – MOC SOFTWARE INTEGRATION

The MOC software prototype used for this project is based on the one provided by Dr. Jack Chosnek of Knowledge One process safety management consulting firm. The MOC form is adapted from Garland[27]. The software is built using File Maker Pro. A simple flowchart of the method is shown in Figure 18.

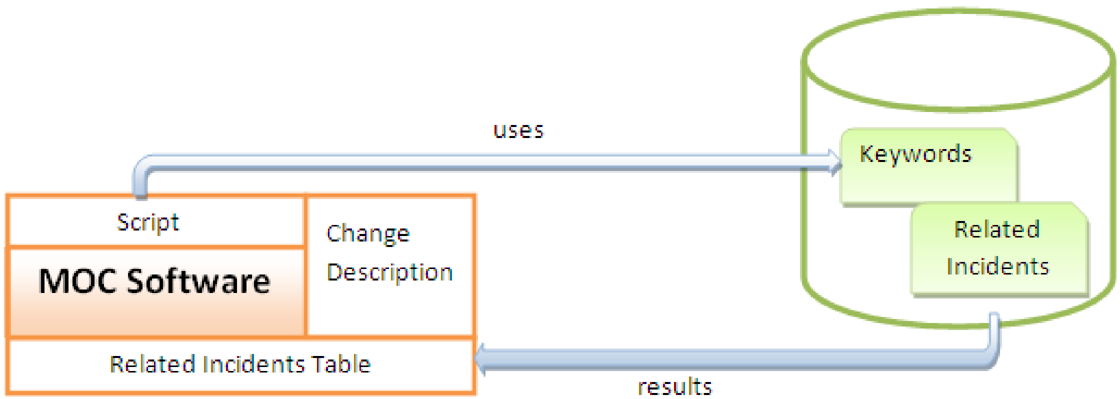

Figure 18. Flowchart depicting how the active database integrated with the MOC software would return result

The development process is summarized below:

- A short list of keywords has been prepared for a few related incidents from the MARS database.
- This information was used for populating a database table (keywords table)
- A prototype MOC application (MOC form) was created using FileMaker Pro 10.0v1.
- A FileMaker script was created to run on the MOC form. The script reads the description of change in the MOC interface and looks for matching keywords in the 'keywords table'. If there is a hit, it displays the related incidents.

- The user can click on the incidents to open a pop up window containing the incident report.

Figure 19 shows the screenshot of the database application on MOC. The script pauses once the first keyword is highlighted and gives the user an option to click on it to reveal a list of related incidents, shown as file numbers in the right bottom corner of the screen.

A click on any of these links will generate a pop up window with the related incident information as shown in Figure 20. Then to re-run the script to catch the next keyword, the user can click on the continue button on the top right corner of Figure 19.

The keywords used for the active system is limited to hazardous chemicals and equipment list. Whenever a single keyword match is found with the incident files in the database, the system returns the first ten results. This can be enhanced, for example by using two or multiple keywords match, which may ensure better results are ranked on the top and thus fall under the top ten results that are displayed in the MOC screen. A fuzzy search criterion can also be employed to have better set of results.

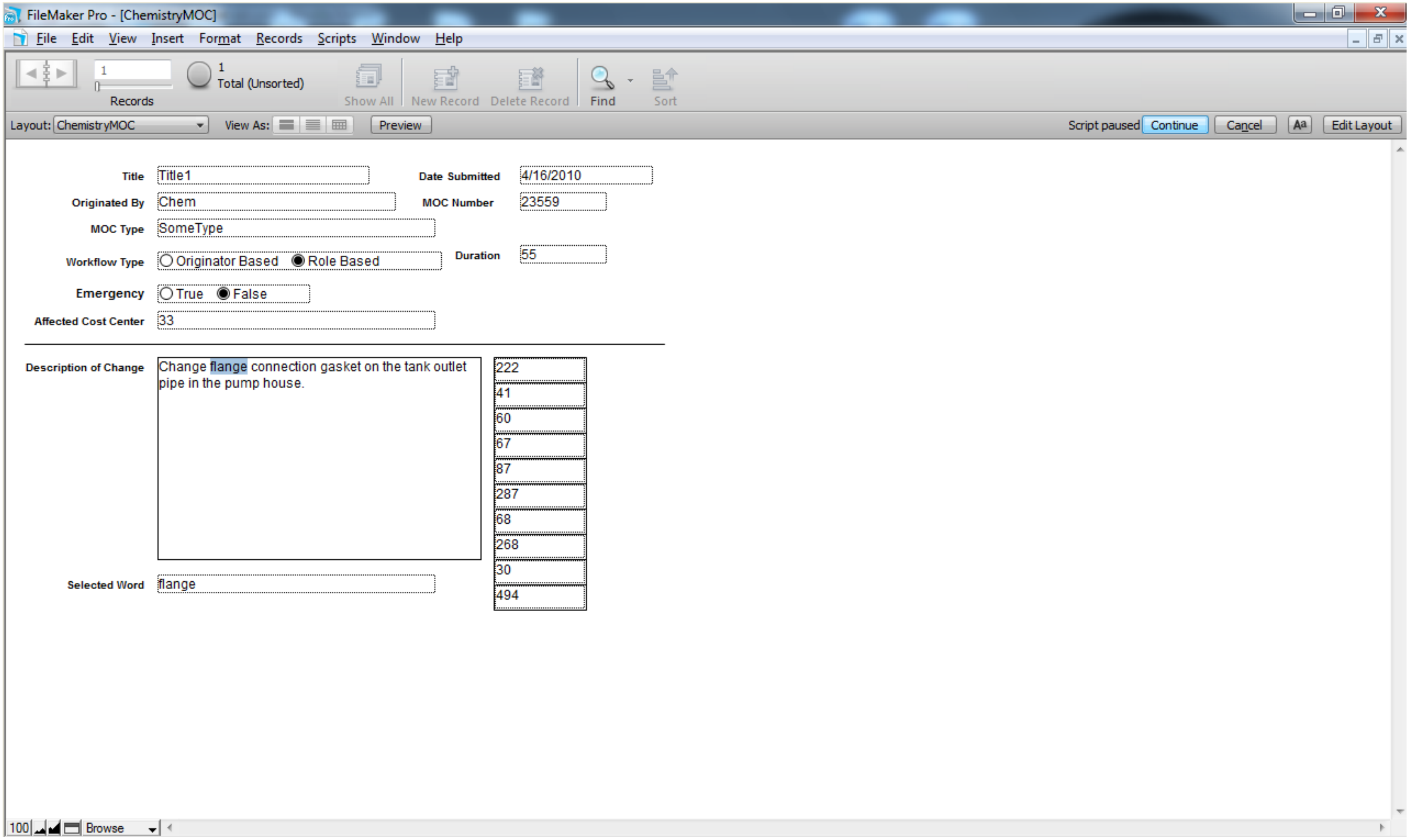

Figure 19. Screenshot of MOC form in the program prototype

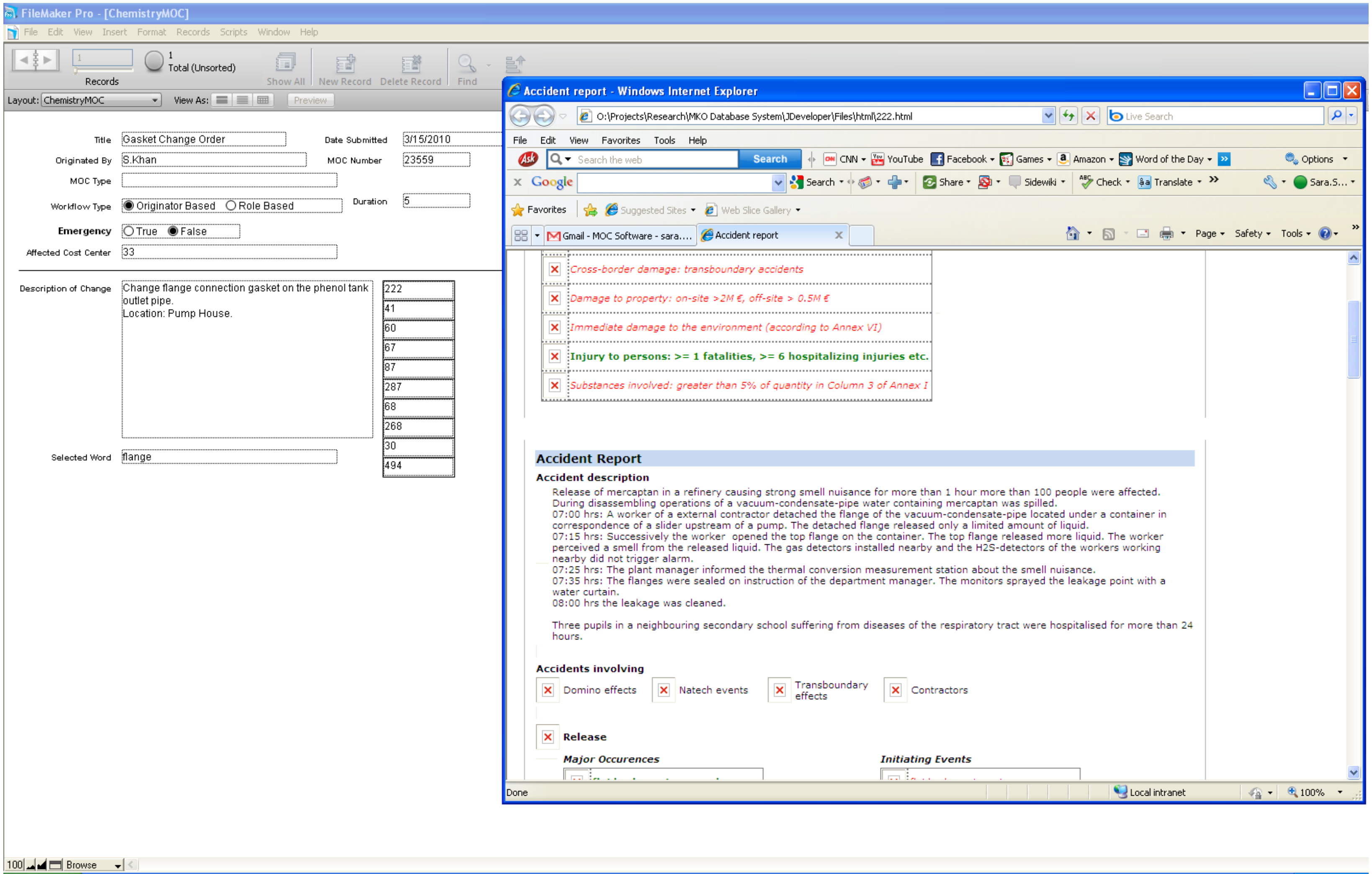

Figure 20. Screenshot of the pop up window when one of the incident file links is clicked

8. CONCLUSIONS AND FUTURE WORK

In this research text mining techniques on MARS database and domain knowledge of process safety has been utilized to develop an incident taxonomy. The taxonomy was developed to serve as the hierarchy for an incident folder search system. Searching incident information using folder search as well as word search can greatly narrow down search results relevant to user needs. In addition, the idea of an active search system has been explored through integration of database with an MOC software prototype.

Incident databases contain unstructured data, which can provide useful information about root causes and lessons learned. Text mining is a useful tool for analyzing large set of unstructured data by converting the text to structured data, which can then be used for further statistical or data mining analysis. Text mining applications on MARS database provided some useful insight on relationship between different keywords, which are helpful for building an incident taxonomy. However, the basic structure of the taxonomy had to come from process safety domain knowledge through literature review. Indeed text mining is only a tool that need educated guidance from domain knowledge to return the desired results. There is however, additional room for improvement on the text mining results through:

- Further manipulation of the stop word list, start word list, phrases, and synonyms list until the total number of words are reduced to a minimum, consisting of non-redundant useful keywords;
- Starting with an educated list of keywords that conforms to the domain (process safety); this initial list can then be reduced by pairing up the terms that have the same connotations. However, this is an iterative step and there needs to be a stopping point, which is easier if the desired results are achieved within a reasonable timeframe.
- Narrowing down the data scope by choosing one type of data; for example using a single variable may return more results within a narrow scope, which is easier

to analyze. In text mining, a large number of smaller text data is preferable as opposed to a fewer number of large text data. This is because large data tend to contain numerous terms and hence no term can characterize individual data.

The taxonomy prepared for this research is a ‘first pass’ of incident taxonomy dedicated for categorizing keywords, which can help users find incident information they need but do not know how to ask for. There is still room for improvement and modifications. However, care should be taken as to keeping the purpose and goal in mind. For incident database searching purpose, the taxonomy needs to be kept as simple as possible, so that both tasks of finding and inputting information are easy. Otherwise people will not be motivated to use such system.

The database arranged in a taxonomy is implemented in Microsoft Excel format, although initial attempts were made to use an open source Content Management System (CMS) for the purpose. However the database system created for this research is expandable and can be transformed to a CMS in future. It is highly recommended to use a CMS to store the database, because of the ease of managing a large set of data. The folders created in a CMS are virtual folders, which can be linked with other folders by pointers. Thus, it provides more flexibility in updating new information and changing the structure of the taxonomy with the flow, according to changing user/administrator needs. Once a fully fledged incident database is developed, it is very important for keeping the valuable data secured. This can be done by giving administrative privilege to one or two persons responsible for updating the data. Also maintaining a single data reporting format is useful for interpreting the information required or conveyed by both parties – reporter and user. Finally, a web-based system with user login capabilities will enhance the usability of the database system and will ensure growth.

The active search system as designed for this research uses keywords to find related incidents. At this point, a single keyword match with an incident file is returning that file as result, if it is among the top ten results. This can be enhanced, for example by using two or multiple keywords match, which may ensure better results are ranked on the top and thus fall under the top ten results that are displayed in the MOC screen. In addition, a set of intelligent keywords based on knowledge of process safety can ensure better matches and thus better search results.

An active and knowledge-based incident information retrieval system can enhance our learning from past mistakes and make us more vigilant to incident records before making an educated, safer decision.

LITERATURE CITED

1. Kletz, T. A., HAZOP & HAZAN notes on the identification and assessment of hazards. In *IChemE Rugby,* Warwickshire, UK, **1983**.

2. Chung, P. W., Jefferson, M., The integration of accident databases with computer tools in the chemical industry. *Computers Chem Engng* **1998,** *22*, 729-732.

3. Sepeda, A. L., Lessons learned from process incident databases and the process safety incident database (PSID) approach by the Center for Chemical Process Safety. *Journal of Hazardous Materials* **2005,** *130*, 9-14.

4. CDC-ATSDR Chemical emergencies fact sheet: Hazardous Substances Emergency Events Surveillance. http://www.atsdr.cdc.gov/hs/hsees/hsees_about-factsheet.pdf (accessed Jan 3, 2010).

5. EPA-RMP Risk Management Plan rule. http://www.epa.gov/oem/content/rmp/index.htm (accessed Jan 3, 2009).

6. NRC National Response Center. http://www.nrc.uscg.mil/nrcback.html (accessed Oct 5, 2009).

7. OSHA Occupational Safety and Health Administration. http://www.osha.gov (accessed June 3, 2008).

8. U. G. A. Office, *Aviation Safety: Undeclared air shipments of dangerous goods and DOT's enforcement approach*, 2003. http://www.gao.gov/ (accessed Jan 1, 2010).

9. DOT-PHMSA US Department of Transportation Pipeline and Hazardous Materials Safety Administration. http://www.phmsa.dot.gov/ (accessed Feb 15, 2010).

10. MAHB-MARS Major Accident Reporting System. http://mahbsrv.jrc.it/mars/default.html (accessed Jan 09, 2008).

11. Chung, P. W. H., Jefferson, M., A fuzzy approach to accessing accident databases. *Applied Intelligence* **1998,** *9*, 129-137.

12. Anand, S. Novel applications of data mining methodologies to incident databases. M.S. thesis, Texas A&M University, College Station, TX, 2005.

13. OSHA, A Guide to *The Globally Harmonized System of Classification and Labeling of Chemicals* (GHS). http://www.osha.gov/dsg/hazcom/ghs.html (accessed Mar 15, 2010)

14. Markowski, A. S., *Layer of Protection Analysis for the Process Industries*. Polska Akademia Nauk: Lodz, 2006.

15. Dianous, V., Fiévez, C., A more explicit demonstration of risk control through the use of bow-tie diagrams and the evaluation of safety barrier performance. *Journal of Hazardous Materials* **2006,** *130*, 220-233.

16. Hale, A. R., Ale, B.J.M., Goossens, L.H.J., Heijer, T., Bellamy, L.J., Mud, M.L., Roelen, A., Baksteen, H., Post, J., Papazoglou, I.A., Bloemhoff, A., Oh, J.I.H. , Modeling accidents for prioritizing prevention. *Reliability Engineering and System Safety* **2007,** *92*, 1701-1715.

17. Delvosalle, C., Fiévez, C., Pipart, A. *ARAMIS Project: Deliverable D1C-WP1*; Faculté Polytechnique de Mons, European Commission: Mons, France, 2004.

18. Bond, J., C-MIST, Linking an accident database to design and operational software. *IChemE Symposium, Hazards XVII*, Manchester, UK, 2003.

19. Kletz, T., Searchlights from the past. In *9th Annual Symposium, Mary Kay O'Connor Process Safety Center*, Texas A&M University, College Station, TX, 2006.

20. Nisbet, R., Elder, J., Miner, G., *Handbook of Statistical Analysis and Data Mining Applications*. Elsevier: Burlington, 2009.

21. Kirchsteiger, C., EC DG-JRC, MAHB, Technical Guideline on Reporting Accidents to the MARS Database. European Commission: Ispra (VA), Italy, 2001.

22. Raja, U., Tretter, M.J., Classification of software patches: a text mining approach. *Journal of Software Maintenance and Evaluation: Research and Practice* **2010**, (accepted).

23. Statsoft electronic statistics textbook. http://www.statsoft.com/textbook/ (accessed Jan 3, 2009).

24. Abbasi, T., Pasman, H.J., Abbasi, S.A., A scheme for the classification of explosions in the chemical process industry. *Journal of Hazardous Materials* **2010,** *174*, 270-280.

25. Sklet, S., Safety barriers: definition, classification, and performance. *Journal of Loss Prevention in the Process Industries* **2006,** *19*, 494-506.

26. Apache Lucene. http://lucene.apache.org/java/docs/ (accessed Mar 7, 2010).

27. Garland, R. W., Electronic management of change process. *Process Safety Progress* **2004,** *23*, (4), 244-251.

APPENDIX A

A list of the initial start word list used for text mining all text variables of MARS database short reports is shown below.

Absorption	Detector	HCL	Monitor	Secure
Addition	Detonate	Head	Motor	Sensor
Air	Develop	Health	Naphtha	Separate
Alarm	Diesel	heat	Near	Severe
Alert	Dilute	High	Nitrate	Sewage
Ambient	Dimension	History	Nitric acid	Shell
Ammonia	Dioxide	Hit	Nitrogen	Shelter
Ammonium	Discharge	Hole	Noise	Shift
Atmosphere	Disconnect	Hose	Off-site	Ship
Audit	Dismantle	Hospital	Oil	Shock
Authorised	Disperse	Hot	On-site	Shut
Automatic	Disrupted	House	Onsite emergency plan	Shutdown
Barrier	Distance	Human	Open	Signal
Basin	Distilled	hydraulic	Operate	Skin
Batch	Door	Hydrocarbon	Organs	Smell
Benzene	Double containment	Hydrochloric acid	Outlet	Smoke
Blast	Downstream	Hydrogen chloride	Overflow	Sodium
Block	Drain	Hypochlorite	Overheated	Soil
Blown	Drop	Ignite	Overpressure	Solid
Boiler	Drum	Ignition source	Oxide	Solution
Bolt	Dry	Immediate	Oxygen	Solvent
Bottom	Dust	Impact	Permit to work	Source
Break	Ecologic	Increase	Pesticide	Spark
Brigade	Electric	Inert	Petrochemical	Speed

Build	Eliminate	Initiating event	Pharmaceutical	Spill
Bund	Emerge	Inject	Phase	Spontaneous
Burn	Emission	Injury	Phenol	Spray
Burst	Employee	Inlet	Phenolproof	Spread
Butane	Empty	Inspect	Pipe	Sprinkler
Cable	Energy	Install	Pipeline	Start
Calculate	Engine	Instrument	Pit	Start-up
Capacity	Environment	Insulated	Plastic	State
Car	Equip	Integrity	Plate	Steam
Carbon	Escape	Interrupt	Plug	Steel
Carbon monoxide	Ethylene	Intervene	Police	Store
Catalyst	Euro	intervention	Pollute	Structure
Catchment	Evacuate	Intoxicated	Polyethylene	Submerged
Chamber	Evaporate	Inventory	Polymeric	Substitute
Change	Exceed	investigate	Pool	Suction
Channel	Excess	Iron	Populated	Sulphur
Charge	Exhaust	Irritant	Powder	Sulphuric acid
Check	Exotherm	Isolate	Power	Supply
Chemicals	Expert	Issue	Precipitate	Surface
Chlorine	Explosion	Jet	Pressure	Suspension
Circuit	Exposed	Kill	Pressurised	Switch
Classification	External	Knowledge	Prevent	Synthesis
Clean	Extinguish	Laboratory	Procedure	Tank
Close	Extract	lack	Produce	Tanker
Cloud	Eye	Land	Prohibited	Temperature
Coat	Face	Layer	Propagate	Test
Collapse	Fail	Leak	Propane	Thermal
Column	Far	length	Protect	Threshold
Combusts	Fast	lesson	Pump	Toluene

Compliance	Faulty	Level	Purge	Tower
Component	Fauna	Life	Railway	Toxic
Compound	Feed	Lift	Rain	Train
Compressed	Fill	Line	Range	Transfer
Compressor	Filter	Liquid	Rate	Transport
Concentrate	Fire brigade	Live	React	Treatment
Concrete	Fire-brigade	Load	Reaction	Truck
Condense	Fire-fighter	Lost	Reactor	Tube
Confine	Fire water	Low	Recurrence	Underground
Connect	Fish	Lpg	Recycle	Unload
Consequence	Fix	Machine	Reduction	Upstream
Contamined	Flame	Main	Refinery	Use
Content	Flammable	Maintain	Refrigerant	Vacuum
Control	Flammable liquid	Maintenance	Regulate	Valve
Cool	Flange	Malfunction	Release	Vapour
Corrosion	Flare	Man	Relief	Vapour cloud
Cost	Flash	Manager	Remedial	Vehicle
Cover	flexible	Manual	Remote	Vent
Crack	Flora	Manufacture	Repair	Ventilated
Crude	Flow	Material	Replace	Vessel
Curtain	Fluid	Maximum	Rescue	Violent
Cut	Flush	Measure	Residue	Wall
Cylinder	Foam	Mechanic	Resin	Warehouse
Data	Form	Medical	Respiratory	Warn
Death	Fragment	Melt	Respond	Waste
Decomposition	Fuel	Metal	Restart	Water
Decontaminate	Full	Methane	Reveal	Water curtain
Defect	Fume	Methanol	Risk	Weather
Deflagrant	Function	Methyl	River	Weld

Deform	Furnace	Miss	Road	Well
Degraded	Gas	Mistake	Roof	Wind
Density	Gasket	Mitigate	Rupture	Window
Depot	Gasoline	Mix	Safe	Wrong
Design	Glass	Mixer	School	
Destroy	Good	Mixture	Screw	
Detect	Ground	Modify	Seal	

The following is a list of the stop words used for text mining all text variables in addition to the default English stoplist of Statistica.

abate	attached	done	may	remarks
able	attempt	drawn	mean	represent
abnormal	attend	due	meet	request
absence	attention	either	mention	requested
access	august	effect	mentioned	respect
accident	authors	enough	metres	result
accidents	avail	ensure	minute	resulted
accord	Away	establishmentaffected	minutes	resulting
according	based	etc	might	second
account	basis	even	mm	sent
across	became	event	moment	seveso
act	beside	eventual	month	should
action	besides	every	morning	side
actual	brought	extent	much	significant
ad	ca	fact	must	since
adjoins	can	february	name	situated
administrative	carry	first	necessary	six
adopt	cas	five	night	slight
advice	category	found	north	soon
advise	caught	four	now	south
advised	cause	furthermore	obtain	special
affect	causes	gave	obtained	specific
affected	caused	general	occur	stay
afterwards	causing	give	occurred	subject

agency
agent
ago
aid
allow
allowed
allowing
almost
along
already
also
although
among
annex
another
anymore
apart
apparent
apparently
appear
applicable
application
appointed
appropriate
approx
approximate
approximately
area
around
arrange
arrival
arrive
arrived
article
ask
aspect
assist
associated
assume
assure
centre
certain
characteristic
characteristics
chemical
circumstance
cm
code
company
companies
comprise
comprises
concern
Concerned
concerning
consequent
consider
Considerable
Considered
considering
considers
consist
consult
Correspond
corresponding
could
course
criteria
date
day
de
decide
decided
decision
declare
describe
described
description
despite
determine
given
go
going
got
happen
happened
hour
however
hrs
ii
incide
incident
include
Included
Including
industrie
industry
initial
instead
interest
involv
Involve
Involved
involving
just
keep
kept
kind
last
lead
led
left
like
longer
kg
litre
m2
m3
make
made
occurrence
onto
one
order
ordered
particular
particularly
per
period
place
placed
places
possible
possibly
prefect
prepare
Prepared
prepares
Present
Presented
probable
probably
Provide
provided
purpose
put
question
rd
Reach
reached
realise
reason
Receive
received
Refer
reference
regard
Relate
relates
relevant
subsequent
subsequently
summary
suppose
take
taken
taking
team
th
therefore
third
though
three
Thus
tonne
took
total
toward
try
two
Type
undertaken
various
via
way
wasn
week
went
west
whereas
whether
whilst
will
Within
without
would
year
yes
yet

assured	didn	manage	remain
attache	discuss	managed	remains

Next is a list of the phrases used for all text variable text mining analysis.

ammonium nitrate
Back pressure
block valve
Boiling liquid expanding vapour explosion
boiling point
carbon dioxide
carbon monoxide
cellulose nitrate
check valve
Comity of competent authorities
Committee of competent authorities
control valve
distributed control system
double containment
ethylene oxide
fire ball
fire fighter
fire water
flammable gas
flammable liquid
Flammable mixture
flammable vapour
flash fire
flash point
flexible hose
float valve
gas cloud
hazard analysis
hazard identification
hydrochloric acid
hydrogen chloride
hydrogen fluoride
ignition source
operating condition
operating temperature
operating pressure
permit to work
Personal protective equipment
Personal protective gear
polymeric isocyanate
polymeric resins
polymeric SO3
pressure wave
process hazard analysis
process safety
quantitative risk assessment
rail tank
railway tanker
recycle gas
reduction reaction
relief valve
retention basins
risk asessment
runaway reaction
safety culture
safety measure
sample point
set pressure
set temperature
shock wave
sodium cyanide
sodium hypochlorite
spontaneous ignition
spontaneous decomposition
spontaneous polymerisation
standard operating procedure

initiating event
management of change
melting point
methyl chloride
methyl hydroperoxide (MHP)
methyl pyridylketone
natural gas
nitric acid
nitrogen dioxide
onsite emergency plan
sulphuric acid
titanium dioxide
transfer tank
trichloro (methyl) silane
waste water
water curtain
work permit
vapour cloud
vapour cloud explosion

Finally, the declared synonyms for text analysis of all variables are listed below.

difficult: difficulty
dimension: diameter, length
discover: discovered
dismantle: disconnect
disperse: spread, spreading, dispersed
disrupted: disturbed
distributed control system: DCS
drain: drainage
electric: electrical
emission: emitted
employee: operator, staff, worker, personnel, crew
empty: emptied
environment: environmental, flora, fauna, fish
escape: escaped, escaping
establish: established, establishment
euro: euros
evacuate: evacuation, evacuated
evaluate: evaluated
examine: examined
excess: exceed
experience: experienced, expert
explosion: explode, exploded
exposed: exposure
extend: extension
extinguish: extinguished, extinguishing
facility: facilities
fail: failure, failed

fall: fell
feed: fed, feeding
fill: filled, filling
fire-fighter: firefighter, fire fighter, firemen, fire-fighting, fire fighting, fire-brigade, fire brigade
follow: followed, following
form: formed, forming
full: fully
gas: gases, gaseous
good: better, great
hazard: dangerous, danger, harm
hcl: hydrochloric, acid, hydrogen, chloride
heat: heated, heating, hot
high: higher, height
hold: held
hospitalized: hospitalised
house: home
identify: identification, identified
ignite: ignited
immediate: immediately
implement: implemented
improve: improved
inadequate: insufficient, lack
increase: increased, rise
indicate: indicating, indicated, indication
inform: informed
initiate: initiated, initiating
injury: injured, injures, injuries
inspect: inspected, inspector
install: installed
interrupt: interrupted
intervene: intervention, interven, intervened
inventory: inventories
issue: problem
knowledge: lesson, learn
leak: leakage, leaked, leaking
located: located
loss: lost
low: lower
lpg: liquefied, petroleum, gas
maintain: maintained, maintenance

man: men
manager: supervisor, supervise
manufacture: manufacturing
maximum: max
measure: measured
methane: methan
methylethylketon: MEK
mistake: error, wrong
mix: mixed, mixing, combined, mixture
monitor: observe, observes, observed
near: nearby, vicinity, proximity, adjacent, nearest
need: needed
neighbour: neighbourhood, neighboring
normal: normally
note: noted
note: notice
off-site: offsite
on-site: on-sit, onsite
open: opened
operate: operated, operating
originate: originated, originating
overpressure: over-pressure, over pressure
partially: partly
pass: passed
people: person, populated, public, human
perform: performed, performing
pipe: pipework
plan: planned
pollute: polluted, pollution
ppm: parts per million
precaution: precautionary
prevent: prevented, eliminate, avoid, precaution
previous: prior, previously
process: processing
produce: produced, production, producing
program: programme
prohibited: restrict
protect: protected
pump: pumped, pumping
quality: qualities

quantity: quantities
quick: rapid, quickly, rapidly
recover: recovery, recovered
recycle gas: ricycle gas
reduce: reduced, reduction
region: regional
release: discharge, discharged, emission
remove: removed, removal
repair: repaired
replace: replaced, substitute
report: reported
require: requires, required, requirement
resident: residential, house
residue: residual, accumulate, deposit
respond: response
return: returned
review: reviewed
run: running
rupture: ruptured
safe: safety
sample point: sample points
seal: sealed, sealing
see: seen
separate: separated, separation
severe: serious, intense, violent
show: shown, showed
shutdown: shut-down, shut-off
spill: spillage
spontaneous decomposition: decomposed spontaneously
spontaneous ignition: ignited spontaneously
start: started, starting, begin, began
stop: stopped
store: storage, stored
structure: structural
supplier: suppler
supply: supplies, supplied
temporary: temporarily
test: tested, testing
tetrahydrofurane: THF
time: times

ton: tonnes, tonne
training: train
transfer: transfers, transferred
transport: transports, transported
treatment: treat
use: used, using
vapour: vapor
vapour cloud: gas cloud
vent: venting
warning: warn, warned
waste: effluent
work: worked, working, duty
work permit: permit to work, permit-to-work

A list of the start words chosen for text mining the causes variable data is shown below.

Pre startup safety review
Procedures
Process upset
Pump
Rail
Rain
Reactive chemical
Refractory
Release
Repair
Resistance
Reverse flow
Risk analysis
Road
Roll-over
Runaway reaction
Rupture
Rust
Sabotage
Safety equipment
Safety management
Safety relief
Sampling
Seal
Shear stress
Shift change
Shipping
Short-circuit
Side reaction
Slip
Smoking
Snow
Solid
Soluble
Spark
Spontaneous combustion
Stabilizer
Static electricity
Steam
Steel
Storage
Storm
Stress
Cracking
Tank
Temperature
Tensile
Terrorism
Secomposition
Thermal
Expansion
Weakening
Thermite reaction
Tool
Toxic
Training
Transportation
Tube
Uncontrolled
Underfilling
Underpressure
Unstable
Utility
Valve
Vandalism
Vapour
Vapour cloud
Vessel
Vibration
Violation
War
Water

Shaft	Testing	Wind

Following is a list of the causes phrases:

accidental mixing
air system
air vector
arc welding
auto ignition
back pressure
back flow
bursting disc
chemical reaction
computer system
contributing factor
cooling tower
cost cutting
domino effect
dust explosion
earth movement
electric installation
emergency shutdown
emergency shut-down
emergency shut down
equipment failure
exhaust gas
exothermic reaction
external force: External stress
external overpressure
fast emptying
flexible coupling
flexible hose
floating roof
thermal decomposition
thermal expansion
thermal weakening
thermite reaction
unstable solid
vapour cloud release
flow rate
fuel supply
gas flame welding
hazard analysis
hazardous substance
hazardous substances
hazardous material
high pressure
high temperature
hot missile
human factors
hydrogen embitterment
incompatible reagent
initiating event
Instrumentation failure
Insulating plate
lack of earthing
lock out
low pressure
low temperature
magnifying glass
management of change
material of construction
operation
organic peroxide
oxidizing substance
oxygen enrichment
permit to work
plane crash
power supply
pre startup safety review
process upset
reactive chemical
relief valve
reverse flow
risk analysis
roll over
runaway reaction
safety equipment
safety management
safety relief
safety relief valve
safety valve
shear stress
shift change
side reaction
spontaneous combustion
static electricity
steel grade
stress corrosion cracking

Following is a list of the causes synonyms:

mix: mixed, mixing
age: aged, aging

assemble: assembling, assembled
autoignition: auto ignition
backflow: back flow
block: blocked, blocking
catalyst: catalytic
clean: cleaning, cleaned
code: codes, standard, standards, rules, regulations
combustible: combustibility, combustion, combusted
communication: communicated, communicate, information flow
compress: compressed, compression, compressing, compressive
condensation: condense, condensed, condensing
conductor: conductivity, conduct
deficiency: lack, absence, insufficient, inadequate
degradation: degrade, degraded, damage, damaged
design: designing, designed
detection: detector, detect
disconnect: disconnected, connection failure
documentation: document, documents
drain: draining, drainage
earthquake: earth movement
electric: electrical, electric, installation
ergonomic: ergonomics
expansion: expand, expanded
explosive: explosivity
fragment: fragmented
freeze: freezing
frothing: froth
implementation: implement, implemented, implementing
impurities: impurity
inspection: inspect, inspected
instruction: instructed, instruct
isolation: isolate
labelling: label, labels, labelled
leak: leaking, leaked, leaks
lifting: lift, lifted
lockout: lock, out, lock-out
lubrication: lubricate, lubricated, lubricating
manning: manned, man
misalignment: not aligned
mistake: incorrect, wrong, error

omission: omit, omitted
open: opened, opening, opens
operation: operate, operating, operator
organizational: management
overfilling: overfill, overfilled
overflow: overflowing, overflowed
overheating: overheat, overheated
overload: overloaded, overloading
planning: plan
policy: policies
polymerization: polymerize
pump: pumping, pumped
rail: railway, rail-way
repair: fix
resistance: resistant
roll-over: rollover, roll over
rupture: breach, ruptured, breached
rust: rusting, rusted
sampling: sample
shipping: ship, shipped
short-circuit: short circuit
storage: store, storing, stored
stress: force
tank: tanks, storage tank
tensile: tension
terrorism: terrorist
testing: test, tested
toxic: poisonous, poison
training: train, trained
transportation: transport
uncontrolled: uncontrollable
underfilling: underfill, underfilled
vibration: vibrate, vibrating, vibrated
violation: violate, violated, violating

APPENDIX B

Table B.1 Cluster Members and Distances for All Variables k-means Cluster Analysis

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
ethylene oxide	1	0.945	16	27	**flammable gas fire with some toxic release**
sulphuric acid	1	0.603	168	368	
boiler	1	0.516	178	337	
oxide	1	0.508	31	44	
oxygen	1	0.466	66	378	
waste water	1	0.431	11	25	
pharmaceutical	1	0.418	12	51	
tube	1	0.391	91	135	
solvent	1	0.390	14	14	
feed	1	0.386	143	239	
onsite emergency plan	1	0.384	7	14	
exhaust	1	0.378	33	94	
hydrogen fluoride	1	0.376	16	30	
carbon monoxide	1	0.349	22	94	
hydrogen chloride	1	0.333	24	79	
polyethylene	1	0.331	18	29	
methanol	1	0.324	17	51	
carbon	1	0.320	9	13	
double	1	0.317	9	20	
catalyst	1	0.312	71	157	
increase	1	0.311	17	24	
sodium hypochlorite	1	0.297	22	89	
thermal	1	0.289	18	28	
start-up	1	0.283	51	72	
flammable liquid	1	0.279	20	73	
dry	1	0.276	22	39	
burst	1	0.273	34	86	
hydrocarbon	1	0.269	12	20	
carbon dioxide	1	0.263	22	38	
shock wave	1	0.262	17	41	
weld	1	0.259	98	140	
hydrochloric acid	1	0.259	25	57	
channel	1	0.252	26	47	
sprinkler	1	0.249	62	84	

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
nitric	1	0.245	62	479	
install	1	0.245	31	41	
defect	1	0.242	89	148	
instrument	1	0.241	130	257	
cylinder	1	0.241	17	37	
ammonium	1	0.240	47	179	
runaway reaction	1	0.240	7	7	
compressed	1	0.240	9	9	
barrier	1	0.239	28	43	
pesticide	1	0.238	22	51	
layer	1	0.238	11	19	
motor	1	0.236	25	110	
employee	1	0.233	56	69	
evacuate	1	0.225	49	95	
jet	1	0.223	300	707	
train	1	0.219	8	15	
toluene	1	0.218	22	30	
hole	1	0.217	59	143	
hit	1	0.217	191	249	
batch	1	0.217	33	58	
control valve	1	0.216	15	20	
remote	1	0.214	33	146	
mixer	1	0.212	10	30	
experience	1	0.211	40	45	
weather	1	0.211	41	125	
inadequate	1	0.209	33	49	
suspension	1	0.207	28	32	
glass	1	0.206	20	61	
fire water	1	0.202	101	136	
signal	1	0.202	26	33	
impact	1	0.202	21	30	
cool	1	0.198	18	28	
plastic	1	0.197	12	16	
knowledge	1	0.197	23	50	
cover	1	0.197	9	13	
dilute	1	0.197	11	16	
load	1	0.195	21	33	

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
relief valve	1	0.195	31	51	
respiratory	1	0.193	53	151	
burn	1	0.187	25	42	
head	1	0.187	13	15	
flash fire	1	0.187	27	50	
medical	1	0.185	42	93	
rescue	1	0.185	40	69	
door	1	0.185	452	1414	
monitor	1	0.185	60	94	
hose	1	0.184	46	73	
school	1	0.184	26	35	
integrity	1	0.183	13	17	
manufacture	1	0.183	141	266	
drum	1	0.182	58	78	
good	1	0.182	32	134	
vacuum	1	0.181	9	81	
audit	1	0.181	22	32	
water curtain	1	0.179	31	38	
energy	1	0.179	20	49	
noise	1	0.179	41	54	
suction	1	0.173	262	1021	
spray	1	0.173	44	57	
inert	1	0.172	258	474	
bolt	1	0.168	38	56	
resin	1	0.167	17	21	
injury	1	0.165	44	85	
methyl	1	0.162	7	10	
block	1	0.161	23	82	
disperse	1	0.161	18	34	
ignition source	1	0.161	465	3318	
cable	1	0.158	34	55	
reduce	1	0.157	127	219	
residue	1	0.157	16	22	
death	1	0.155	22	35	
detect	1	0.150	121	205	
plug	1	0.147	29	80	
absorption	1	0.146	7	10	

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
insulated	1	0.146	12	21	
mistake	1	0.145	7	8	
boiling point	1	0.145	16	25	
flammable gas	1	0.145	95	174	
issue	1	0.145	21	25	
drop	1	0.144	68	180	
intoxicated	1	0.144	37	122	
comity of competent authorities	1	0.144	132	202	
prohibited	1	0.144	111	260	
investigate	1	0.143	35	220	
inlet	1	0.143	37	72	
replace	1	0.142	95	502	
protect	1	0.142	26	95	
hypochlorite	1	0.141	34	55	
consequence	1	0.140	26	30	
overheated	1	0.139	102	184	
shock	1	0.136	119	511	
fluid	1	0.136	31	45	
crack	1	0.135	61	116	
malfunction	1	0.134	74	127	
smell	1	0.133	160	301	
safety culture	1	0.130	111	113	
fill	1	0.128	22	33	
flexible	1	0.128	24	36	
work permit	1	0.128	21	56	
land	1	0.127	81	130	
flash point	1	0.126	18	43	
permit to work	1	0.125	24	39	
committee of competent authorities	1	0.123	21	47	
spark	1	0.121	38	110	
engine	1	0.118	15	57	
live	1	0.118	9	11	
break	1	0.118	125	196	
flash	1	0.118	25	31	
restart	1	0.112	124	153	

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
fast	1	0.111	52	61	
permit	1	0.109	36	59	
separate	1	0.102	60	80	
pressure wave	1	0.094	12	21	
pressurised	1	0.090	77	79	
on-sit	2	0.674	70	74	**explosion (chemical & physical)**
explosion	2	0.655	24	33	
compressor	2	0.500	25	33	
pipeline	2	0.455	27	31	
powder	2	0.432	11	17	
mix	2	0.409	9	10	
air	2	0.389	9	12	
underground	2	0.381	98	127	
roof	2	0.381	15	23	
tower	2	0.377	39	60	
shell	2	0.373	13	18	
bottom	2	0.354	24	73	
empty	2	0.316	18	32	
chamber	2	0.315	9	10	
dust	2	0.311	18	27	
pit	2	0.308	27	36	
measure	2	0.303	28	51	
metal	2	0.301	19	26	
ground	2	0.297	35	43	
atmosphere	2	0.292	114	288	
ignite	2	0.276	214	649	
benzene	2	0.271	83	94	
flame	2	0.245	14	28	
near	2	0.243	129	348	
test	2	0.217	138	378	
electric	2	0.189	7	8	
repair	2	0.178	24	35	
wall	2	0.099	64	84	
reactor	3	0.944	69	109	**flammable liquid release with resulting fire**
hcl	3	0.816	204	412	

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
tank	3	0.683	20	24	
pipe	3	0.647	29	60	
vessel	3	0.630	113	190	
steam	3	0.619	299	306	
lpg	3	0.611	407	785	
phenol	3	0.596	52	72	
react	3	0.560	58	92	
line	3	0.553	59	75	
valve	3	0.484	50	105	
function	3	0.478	24	45	
plan	3	0.476	24	73	
leak	3	0.414	13	20	
liquid	3	0.401	21	27	
level	3	0.394	131	134	
temperature	3	0.394	114	230	
overpressure	3	0.386	14	46	
phenolproof	3	0.380	21	33	
fire-fighter	3	0.350	19	36	
catchment	3	0.349	18	79	
high	3	0.327	32	120	
pump	3	0.318	98	144	
water	3	0.317	14	16	
full	3	0.307	25	102	
manual	3	0.302	81	202	
flange	3	0.289	23	29	
release	3	0.282	316	585	
alarm	3	0.273	168	472	
automatic	3	0.261	81	222	
emerge	3	0.248	17	26	
maintain	3	0.212	117	183	
control	3	0.182	42	80	
waste	4	0.660	25	41	**environmental/off-site consequences due to toxic release**
bund	4	0.650	8	9	
crude	4	0.630	245	676	

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency
drain	4	0.623	30	88
plate	4	0.555	27	110
smoke	4	0.546	93	123
tanker	4	0.544	19	21
shift	4	0.494	7	11
basin	4	0.457	14	24
sewage	4	0.447	126	173
relief	4	0.445	16	37
vent	4	0.436	86	306
road	4	0.422	80	112
isolate	4	0.405	156	572
manager	4	0.403	71	220
filter	4	0.387	15	59
cloud	4	0.384	161	317
toxic	4	0.375	61	81
sodium	4	0.373	26	66
hospital	4	0.368	17	34
health	4	0.368	116	177
environment	4	0.363	17	30
steel	4	0.362	26	53
contain	4	0.361	29	93
shut	4	0.357	53	127
compliance	4	0.347	11	16
rain	4	0.339	15	20
work	4	0.334	325	432
solution	4	0.333	104	158
hydraulic	4	0.330	8	10
threshold	4	0.325	27	85
compound	4	0.309	20	19
river	4	0.308	150	397
close	4	0.304	29	49
excess	4	0.297	19	39
spill	4	0.296	24	32
rate	4	0.296	39	69
flow	4	0.286	119	303
laboratory	4	0.283	84	416
loss	4	0.280	295	311

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
pool	4	0.279	22	25	
resident	4	0.278	72	264	
downstream	4	0.278	392	938	
wind	4	0.268	33	66	
upstream	4	0.267	13	28	
source	4	0.265	24	33	
outlet	4	0.265	69	105	
maximum	4	0.259	8	11	
open	4	0.249	39	71	
inventory	4	0.249	21	33	
prevent	4	0.249	13	35	
check	4	0.243	7	9	
mitigate	4	0.240	25	35	
well	4	0.238	7	10	
procedure	4	0.231	66	180	
solid	4	0.230	74	160	
detector	4	0.223	70	143	
start	4	0.220	15	26	
overflow	4	0.217	103	174	
vapour	4	0.215	9	12	
shutdown	4	0.208	15	29	
range	4	0.205	35	93	
people	4	0.188	34	38	
exposed	4	0.186	29	66	
clean	4	0.180	17	38	
immediate	4	0.138	53	134	
ambient	4	0.137	22	27	
ammonia	5	0.000	15	30	**ammonia incidents**
natural gas	6	1.006	120	189	**flammable gas/vapor explosion**
ethylene	6	0.871	33	46	
refinery	6	0.740	21	29	
foam	6	0.591	26	35	
oil	6	0.581	76	177	
naphtha	6	0.580	45	124	
seal	6	0.541	13	36	

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
gasoline	6	0.505	64	164	
column	6	0.488	33	59	
methan	6	0.462	73	97	
depot	6	0.437	14	16	
fuel	6	0.408	21	35	
propane	6	0.381	20	34	
flare	6	0.374	16	29	
sulphur	6	0.374	51	84	
lift	6	0.361	52	78	
gas cloud	6	0.329	46	127	
butane	6	0.323	68	108	
low	6	0.300	21	40	
design	6	0.293	45	118	
rupture	6	0.292	23	91	
collapse	6	0.290	17	27	
petrochemical	6	0.288	283	1972	
external	6	0.283	38	109	
extinguish	6	0.282	134	254	
flammable	6	0.281	40	63	
railway	6	0.275	27	42	
off-site	6	0.271	30	41	
escape	6	0.266	28	87	
dioxide	6	0.258	22	82	
heat	6	0.257	136	252	
power	6	0.253	14	22	
vapour cloud explosion	6	0.248	40	76	
vapour cloud	6	0.247	28	73	
police	6	0.237	25	33	
history	6	0.226	32	54	
recurrence	6	0.224	218	876	
interven	6	0.223	51	71	
diesel	6	0.212	25	50	
initiating event	6	0.200	7	11	
density	6	0.199	35	103	
material	6	0.194	91	269	
initiate	6	0.184	59	93	
cost	6	0.181	62	203	

Table B.1 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
life	6	0.157	16	42	
chlorine	7	1.149	369	1295	**decomposition /runaway reaction**
furnace	7	0.950	25	35	
blast	7	0.517	20	25	
decomposition	7	0.493	16	34	
nitrogen	7	0.484	102	133	
nitric acid	7	0.454	71	127	
ammonium nitrate	7	0.444	140	318	
iron	7	0.352	8	11	

Table B.2 Cluster Members and Distances for Causes Variable k-means Cluster Analysis

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
air	1	0.479226	45	70	**electrical equipment failure, overpressurisation of air reservoir, and safety system failure causing fire**
electric	1	0.556505	20	36	
fire	1	0.912052	47	55	
open	1	0.472332	25	30	
safety	1	0.568258	36	55	
system	1	0.497075	10	14	
alarm	2	0.565022	10	13	**Alarm failure or lack of operator intervention on alarm resulting in temperature and pressure rise**
high	2	0.369362	7	10	
pressure	2	0.700422	10	12	
supply	2	0.429417	7	7	
temperature	2	0.449345	21	26	
analysis	3	0.468352	31	49	**Inadequate maintenance, design, or inspection causing equipment failure**
corrosion	3	0.665832	22	25	
design	3	0.377722	12	15	
initiating event	3	0.446503	9	11	
inspect	3	0.428363	82	109	
internal	3	0.336308	25	39	
maintenance	3	0.510390	43	73	
material	3	0.656419	8	24	
metal	3	0.602394	27	63	
pipeline	3	0.695816	25	31	
cloud	4	0.679848	27	27	**Accidental mixing, tank/vessel failure causes release of flammable gas**
explosive	4	0.879623	61	69	
flammable	4	0.417143	13	14	
gas	4	0.905925	151	211	
mix	4	0.440814	23	33	
release	4	0.525305	19	23	
tank	4	1.177462	69	99	
vessel	4	0.701452	33	53	
backflow	5	0.357295	44	57	**control system failure & reactive hazard**
control	5	0.442715	7	12	
cool	5	0.665771	17	20	
emerging	5	0.755865	8	10	
gasket	5	0.938920	7	12	
hydrogen	5	0.878108	104	197	
process	5	0.444535	16	24	

Table B2 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
reaction	5	0.633555	9	10	
valve	5	1.087085	10	12	
water	5	1.095180	12	14	
clear	6	0.315940	108	219	**organizational causes and human error**
deficient	6	0.366718	40	50	
disconnect	6	0.529872	56	67	
drain	6	0.590498	7	7	
equipped	6	0.387393	11	14	
hazard	6	0.600484	15	55	
hose	6	0.513843	94	182	
instruct	6	0.333076	40	131	
leak	6	0.661081	13	24	
liquid	6	1.165229	39	52	
mechanic	6	0.511906	13	16	
operate	6	0.528353	9	11	
organizational	6	0.342119	14	27	
procedure	6	0.499175	27	37	
pump	6	0.931242	54	60	
repair	6	0.310390	7	18	
risk	6	0.445464	38	69	
shift	6	0.675865	8	9	
train	6	0.598971	8	8	
work	6	0.546844	64	68	
atmosphere	7	0.484537	63	74	**Combination of process upset, chemical causes, and equipment failure**
block	7	0.287456	11	16	
burst	7	0.200372	8	9	
change	7	0.300843	34	42	
chemical reaction	7	0.226122	16	23	
clean	7	0.300218	15	18	
combusted	7	0.327691	47	62	
concentrated	7	0.295030	51	93	
contain	7	0.223550	12	13	
crack	7	0.947955	88	111	
decompositio n	7	1.078827	60	80	
degrade	7	0.425580	15	21	

Table B2 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
detect	7	0.269935	18	34	
dust	7	0.387008	16	19	
exhaust	7	0.425415	15	24	
exothermic	7	0.360747	7	12	
exothermic reaction	7	0.222496	50	67	
expansion	7	0.329331	77	107	
external	7	0.440200	7	8	
factor	7	0.241760	30	36	
fast	7	0.238910	8	15	
fatigue	7	0.362412	8	12	
flange	7	0.510924	12	15	
flexible	7	0.212530	12	14	
friction	7	0.308855	13	21	
furnace	7	1.103715	8	8	
high pressure	7	0.351277	24	46	
high temperature	7	0.307766	11	27	
hot	7	0.486756	9	13	
human	7	0.351844	8	10	
hydraulic	7	0.759638	86	144	
identification	7	0.278008	84	99	
impact	7	0.220695	93	113	
instrument	7	0.201985	61	110	
isolate	7	0.560800	49	76	
joint	7	0.268130	102	132	
low	7	0.181276	13	17	
mistake	7	0.260251	20	22	
motor	7	0.545168	25	32	
organic	7	0.226285	11	16	
overfill	7	0.427597	10	13	
overflow	7	0.363119	47	54	
overheated	7	0.276506	75	95	
overpressure	7	0.262289	8	13	
oxygen	7	0.494965	18	27	
permit	7	0.238353	9	12	
plate	7	0.795589	13	23	
polymeric	7	0.383224	7	8	

Table B2 (cont.)

Terms	Cluster	Distance to centroid	Documents	Frequency	Cluster Name
power	7	0.158049	7	8	
relief	7	0.265836	9	9	
roof	7	0.380034	10	10	
runaway reaction	7	0.371113	17	29	
rupture	7	0.390059	17	20	
safety valve	7	0.266790	9	12	
seal	7	0.554987	21	28	
shaft	7	0.299936	16	29	
short-circuit	7	0.259944	99	156	
solid	7	0.256733	164	499	
spark	7	0.162238	44	66	
spontaneous	7	0.178206	13	13	
steam	7	0.761780	7	16	
steel	7	0.350964	8	11	
stress	7	0.340351	22	37	
substance	7	0.379015	13	25	
test	7	0.270933	8	8	
tower	7	0.432991	109	299	
toxic	7	0.187937	24	34	
tube	7	0.427555	44	77	
uncontrolled	7	0.148797	102	198	
vapour	7	0.655221	18	26	
weld	7	0.487157	7	7	
wind	7	0.209762	52	72	

APPENDIX C

Table C.1. Initial, detailed taxonomy developed from literature and text mining results.

Chemicals Taxonomy

State	Gas		
	Liquid		
	Solid		
	Two-phase (gas-liquid)		
	Dust		
	Aerosol		
Hazard Classification	Physical Hazards	Explosives	Mass explosion hazard
			Projection hazard
			Fire hazard or minor projection hazard
			No significant hazard
			Very insensitive substances with mass explosion hazard
			Extremely insensitive articles with no mass explosion hazard
		Flammable	
		Combustible	
		Oxidizing	
		Gases Under Pressure	Steam
		Self-Reactive Substances	
		Pyrophoric	
		Self-Heating Substances	
		Organic Peroxides	

		Corrosive to Metals	
		Emit flammable gases when in contact with water/Dangerous when Wet Material	
	Health Hazard	Acute Toxicity/Poisonous Material	
		Skin Corrosion/Irritation	
		Serious Eye Damage/Eye Irritation	
		Respiratory or Skin Sensitization	
		Germ Cell Mutagenicity	
		Carcinogenicity	
		Reproductive Toxicology	
		Target Organ Systemic Toxicity – Single Exposure	
		Target Organ Systemic Toxicity – Repeated Exposure	
		Aspiration Toxicity	
	Environmental Hazards	Hazardous to the Aquatic Environment	
			Acute aquatic toxicity
			Chronic aquatic toxicity
		Bioaccumulation potential	
		Rapid degradability	
Type	Chemical	Acids	
		Bases	
		Ammonia	
		Chlorine	
		Formulations	
		Hetero-Organics	

		Hydrocarbons	Crude Oil
			Natural Gas
			LPG
			Gasoline, diesel, fuel oil or other petroleum product which is liq@ ambient conditions
		Inorganic compounds	
		Oxy-Organics	Oxidizer
			Organic Peroxide
		Paints and Dyes	
		Polychlorinated Biphenyls	
		Polymers	
		Volatile Organic Compounds	
		Mixtures	
Use	Pesticides/Agricultural	Organochlorine compound	
		Organophosphorous compound	
		Pyrethrin	
		Pyrethroid	
		Dipyridyl compound	
		Chlorophenoxy compound	
		Triazines	
		Thiocarbamates	
		Organo-metallic compound	
		Coumarins	
		Indandiones	
		Convulsants	
		Microbial	
		Dithiocarbamates	
		AChE inhibitors	

		Unidentified cholinesterase inhibitor	
		Includes one or more active ingredients	
	Radiological	Radioactive Materials	
	Biological	Toxic Materials	
		Infectious substance	
	Medical	Magnetized Material	
		Hazardous Waste	

Equipment Taxonomy

Storage Equipment	Storage Tank	Atmospheric Storage	Liquid storage	Single-containment		
				Tank with protective outershell		
				Double containment		
				Full containment (liquid-vapor)		
				Membrane		
				In-ground tank		
				Mounded tank		
				Floating Roof Tank	Cone-roof	
					Dome-roof	
					Dry-seal	
			Solid storage	Silo		
				Hopper		
		Pressurized Tank	Bullet Tank			
			Storage Sphere			
		Refrigerated Storage	Cryogenic tank			
	Container Storage	Cylinder (compressed gas)				
		Drum/barrel				
		Bag				
		Sampling container				
	Storage Depot	Pallets				
		Big bags				
	Salt Dome Storage					
	Underground					

	storage					
Process Equipment	Reactors and Reaction Equipment	Reactors	Non-pressurized	Batch		
				Semi-batch		
				Continuous		
			Pressurized	Batch	Autoclave	
				Semi-batch		
				Continuous	Reforming Equipment	Catalytic Reformer
						Reformer
					Cracking Equipment	Catalytic Cracker
						Cracking Furnace
						Cracking Tower
						Hydrocracker
						Fluid cracker
		Reactor component	Agitators and Stirrers	Agitator		
				Blender		
				Mixer		
	Heating and Cooling Equipment	Cooling Equipment	Air Cooler			
			Cooler			
			Cooling Tower			
			Cooling water system			
			Cryogenic equipment			

			Refrigeration unit	
			Fin fan cooler	
		Heating Equipment	Concentrator	
			Drier	rotary dryers
				spray dryers
				flash dryers
			Reboiler	
			Furnace	
			Incinerator	
			Kiln	
			Oven	
			Heating Coils	
		Heat Exchanger	Shell and tube heat exchanger	
			Plate heat exchanger	
		Phase change equipment	Condenser	
			Crystallizer	
			Evaporator	
			Vaporizer	
	Relief Equipment	Safety Relief Valve		
		Rupture Disc		
		Knockout Drum		
		Header System		
	Separation Equipment	Liquid-liquid separation	Decanting Vessel	
		Fractionating equipment	Distillation Column	Stripper
		Gas-liquid separation	Absorption	

			Scrubber	
		Membrane separation		
		Solid-Gas Separation	Filter	
			Cyclone	
			Precipitator	
		Solid-liquid separation	Adsorption	
			Centrifuge	
			Dessicator	
			Filter	
			Ion exchange bed	
			Settling	
		Solid-solid	Sieve	
	Pressure Raising or Reducing Equipment	Compressor	Strainer	
			Blower	
		Pump	Vacuum compressor	
			Canned pump	Impellor
		Pump or Compressor Components	Pump bearing	
			Pump drive	
			Pump gearbox	
			Pump gland	
			Pump manifold	
			Seal	
		Ejector Equipment		
	Physical Processes Equipment	Mixer		

		Crystallization		
		Adsorption		
		Dryer		
	Packaging Equipment			
	Intermediate process storage equipment	Pressure Vessel		
		Atmospheric vessel		
	Process Control and Instrumentation			
Pipes and Fittings	Supply piping/line	Inlet	Body of Pipe	
			Joint	
			Elbow	
			Tee	
		Outlet	Body of Pipe	
			Joint	
			Elbow	
			Tee	
	Intra-piping	Valve	Drain cock	
			Automatic valve	
			Lever operated valve	
			Manually operated valve	
			Ball valve	
			Bonnet valve	
			Control valve	
			Emergency Shutdown	Fast acting valve
			Foot valve	
			Gate valve	

			Non-return valve
			Petcock valve
			Steam trap
			Safety Relief valve
	Blind or Spade or Slip plate		
	Connector	Bellows	
		Bonnet Joint	
		Compression Fitting	
		Coupling	
		Expansion joint	
		Flange	
		Gasket	
	Drains and Sewers	Drainage System	
		Effluent pond	
		Floating Roof Drain	
		Manhole	
		Sewer	
		Sump	
		Waste Water Treatment	
	Hose		
	Insulation		
	Manifold		
	Nozzle		
	Pipe Support		
Transfer Equipment	Pump		
	Compressor		
	Hoses		

	Gravity flow/chute		
	Mechanical transport	Conveyor	
		Crane	
	Transfer interface	Loading Equipment	
		Unloading Equipment	
Pipeline	Body of Pipe		
	Joint		
	Pipe seam		
	Compressor station		
	Regulator/metering system		
	Scraper trap		
	Bolted fitting		
	Welded fitting		
	Girth weld		
	Sump		
	Repair sleeve		
	Service line		
Electrical Equipment	Transformer		
	Capacitor		
	Generator		
	Battery		
	Cable	Powerline	
	Electrical isolators and switches		
	Motor	Pump-motor	
		Stirrer-motor	
	Drilling Equipment		

	Junction box		
Process Control and Instrumentation	Computer		
	Control box		
	Control cable		
	Control room		
	Instrument or controller	Flow meter or control	
		Level meter or control	
		Process Analyzer	Analyzer designed to detect oxygen
			Analyzer designed to detect moisture
			Analyzer designed to detect acidity, pH
		Pressure meter or control	
		Temperature meter or control	
		Viscosity meter	
		Weighing scale	
Protective and Safety Equipment	Containment doors		
	Emergency equipment		
	Explosion suppression		
	Fire fighting equipment	Fire extinguisher	
		Sprinklers	
	Fire Protection Equipment	Fire/flame detector	
		Fire alarm	
		Smoke alarm	
		Flame arrester	

		Fuseable Link	
	Gas Detector	Flammable gas/vapor detector	
		Toxic gas/vapor detector	
	Overpressure relief	Bursting disc	
		Safety Relief Valve	
		Explosion relief	
		Vacuum relief	
	Personal Protective Equipment	Breathing Apparatus	
		Clothing	
	Safety Shower		
	Vent Systems	Chimney	
		Exhaust	
		Exhaust blower	
		Vent System	
Ancillary Process Equipment	Rail Car		
	Hydraulic System		
	Disposal/Waste Treatment	Incinerator	
		Flare Stack	
		Waste water treatment	
Solids Processing Equipment	Ball mill		
	Chopper		
	Crumbling machine		
	Cutter		
	Extruder		

	Flaker		
	Granulator		
	Grater		
	Grinder		
	Pelletizer		
	Pebble mill		
Mechanical Handling Equipment	Conveyor		
	Chute		
	Fan		
	Hopper feeder		
	Lifting equipment	Crane	
		Elevator	
		Lifting Hoist	
		Winch	
		Wire Rope	
	Winding Machine		
Heating/Cooling for building			
Ventilation System			
Laboratory Equipment			
Tools and Access Equipment	Hand Tools	Hydraulic Jack	
	Power tools	Abrasive wheel	
		Drill	
		Electric sander	
		Grinding wheel	
	Ladder		
	Scaffolding		

Power Plant	Engine		
	Turbine		
	Turbo Generator		
Transport Equipment	Air transport	Aircraft	
		Helicopter	
	Loading/Unloading Point	Loading Arm	
		Ship or Shore Connection	
		Terminal	
	Marine Transport	Container Ship	
		Marine Freighter	
		Marine Tanker	
		Passenger Ship	
	Rails Transport	Freight Train	
		Rail Tanker	
		Rail Wagon	
	River Transport	River Barge	
		River Tanker	
	Road Transport	Bulldozer or JCB or Digger	
		Forklift Truck	
		Lorry	
		Road Tanker	
		Road Vehicle	
	Vehicle fuel tank and associated piping		

Operation Taxonomy

Construction			
Demolition			
Exploration			
Offshore			
Operational Activities	Commissioning		
	Decommissioning		
	Startup		
	Shutdown		
	Emergency		
	Inspection		
	Normal Operations	Draining	
		Purging	
		Sampling	
		Valve Operation	
	Testing		
	Maintenance	Cleaning	
		Entry into Confined Space	
		Hot Work	
		Line Break	
		Preparation for Maintenance	
		Repair	
Processing	Material transfer	Charging Reactor/dosing	
		Loading	
		Unloading	
	Heat transfer	Condensation	

		Cooling	
		Evaporation	
		Heating	
		Vaporization	
	Reaction	Batch Reaction	Catalytic batch reaction
			Thermal batch reaction
		Continuous Reaction	Catalytic continuous reaction
			Thermal continuous reaction
		Polymerization	
	Solids processing	Extrusion	
		Milling	
		Molding	
	Separation	Centrifuging	
		Distillation	
		Filtration	
		Refining	
		Settling	
	Mixing		
	Venting		
Labwork			
Storage			
Warehousing			
Transportation	Rail		
	Road		
	Air		
	Marine		

	River		
	Pipeline		

Causes Taxonomy

Process Upsets	Overpressure	Internal overpressure (Gas material)	Combustion/explosion causes overpressure
			External HP source connected causes overpressure
			Internal high pressure causes overpressure
			Overcompression causes overpressure
			Runaway (side) reaction causes overpressure
		Internal overpressure (liquid)	Overfilling vessel causes overpressure
			Pump causes overpressure
			Runaway (side) reaction causes overpressure
			Thermal expansion of liquid filled vessel causes overpressure
			Back pressure wave causes overpressure
			Combustion/explosion causes overpressure
			Roll-over of vessel contents causes overpressure
			Freezing and expansion of aqueous phase
		Internal overpressure (solid material)	Dust explosion overpressures enclosure
			Unstable solid explodes
	Underpressure	Underpressure (pressure below the containment limit of the vessel)	Chemical reaction consuming gases
			Condensation

			Decrease of pressure due to a decrease of the temperature
			Fast emptying of the vessel
		Underpressure (pressure below the containment limit of the roof)	Condensation
			Decrease of pressure due to a decrease of the temperature
			Chemical action consuming gases
			Fast emptying of the vessel or tank
	Overheating	High temperature	Fire (domino effect)
			Chemical energy (domino effect)
			Furnace, boiler, motor other unshielded hot surface
			Heat transport canalisation
			Resisting electric conductor (resistance, short-circuit)
			Electric arc (defect of the electric installation, arc welding)
			Combustion flame or heat (gas flame welding, smoking)
			Natural event (lightning)
			Hot missile (domino effect)
		Temporary (or permanent) energy source	Fire (domino effect)
			Chemical energy (domino effect)
			Furnace, boiler, motor other unshielded hot surface
			Heat transport canalisation
			Magnifying glass

			Resisting electric conductor (resistance, short-circuit)
			Friction (conveyors, doors, mechanical devices)
			Spark (mechanical)
			Spark (static electricity)
			Electric arc (defect of the electric installation, arc welding)
			Impact
			Combustion flame or heat (gas flame welding, smoking)
			Natural event (lightning)
			Hot missile (domino effect)
	Overfilling	Overflow	
	Underfilling		
	Incorrect flow	Incorrect flow rate	Flow rate too high
			Flow rate too low
		Backflow	
		Overflow	
		Reverse flow	
	Incorrect temperature	High temperature	
		Low temperature	
	Exothermic reaction		
	Flameout		
	Frothing		
	Leak	Air leaking into system	
		Flange leak	
		Gasket leak	
		Joint leak	

		Relief valve Leak	
	Low level of catalyst		
	Roll-over stratified liquid layers (due to temperature/density differences)	Vapor cloud release	
Chemical Causes	Accidental mixing/Contamination	Contact with a combustible (reducing) substance	Presence of an incompatible reagent in the storage room
			Leak, entrainment from a nearby pipe or vessel (domino effect)
			Accidental contamination
			The reducing substances is part of the stored/used chemical (organic peroxide)
		Contact with an incompatible reagent	Presence of an incompatible reagent in the storage room
			Leak, entrainment from a nearby pipe or vessel (domino effect)
			Contamination by natural elements of the environment (air, water)
			Accidental contamination
		Presence of a catalyst (if involved)	Abnormal presence of a catalyst
			Leak, entrainment from a nearby pipe or vessel (domino effect)
			Accidental contamination
		Presence of air vector	Aeration
			Wind
			Gas escaping from a breach or leak (domino effect)

			Air projection due to fall, collapse or explosion of a nearby equipment (domino effect)
			Exhaust gases
		Presence of liquid vector	Liquid escaping from a leak or breach
			Extinction water
			Natural event
			Other water sources
	Oxygen Enrichment	Presence of oxidizing substance	Normal storage/use
			Abnormal storage/use
			Leak, entrainment from a nearby equipment (domino effect)
			Other accidental contamination
	Potentially mobile material	Fragmented material	
		Soluble material	
		Material susceptible to emit vapours (e.g. solvent)	
	Hazardous material storage	Storage of a flammable product	
		Storage of a reaction sensitive chemical	
		Storage of a thermal decomposition sensitive chemical	
	Hazardous material use	Use of a highly reactive substance	
		Use of an explosive material	
		Use of a flammable substance	
		Use of a toxic substance	
	Chemical missing	Low level of catalyst	

		Lack of stabilizer/inhibitor	
		Oxygen deficiency	
	Incompatible material of construction		
	Incorrect chemical concentration		
	Unwanted chemical reaction	Runaway reaction	
		Decomposition	
		Autoignition	
		Spontaneous combustion	Hypergolic liquids mixing
		Thermite reaction	Al + Fe2O3 reaction
		Unwanted polymerization	
Equipment Failure	Material of construction failure	Brittle rupture	Brittle structure
			Impact
		Degradation of the mechanical properties leads to incapacity to sustain high pressure	Corrosion
			Erosion
			Fatigue
			Fracture
			Thermal weakening (lowering of the tensile or compressive stength under the effect of high temperature)
			Other chemical action of the environment (solvent on plastic structure)

			Other physical action on the environment (light on plastics)
			Aging
			Rusting
			Stress
			Embrittlement
		Insufficient Initial mechanical properties of the structure (below normal use conditions)	Inappropriate material
			Inappropriate dimnensions
			Inappropriate assembling
		Excessive external stress	Overloading
			High amplitude vibrations
			Dilation
			External overpressure
			Shear stress
	Instrumentation failure		
	Control failure	Computer system failure	
	Equipment missing	Open manhole	
	Incorrect equipment installed		
	Additional equipment installed	insulating plate	
	Electrical Equipment Failure	Power supply failure	
		Generator failure	
		Motor failure	
		Short circuit	
		Lack of earthing	

		Spark	
		Electric arc (defect of the electric installation, arc welding)	
	Mechanical equipment failure	Accidental opening	Disconnect during filling
			Valve opened by mistake
			Valve left open by mistake
			External force causes accidental opening
			Valve blocked
			Seal, joint loss of effectiveness
			Safety valve, safety relief device
			Fail to clear out contents before opening containment
			Wrong part (containing hazardous material) worked on
			Operation started when containment open
			Uncontrolled flow during sampling/draining
		Leak	Air leaking into system
			Flange leak
			Gasket leak
			Joint leak
			Relief valve Leak
		Blower failure	
		Fittings failure	Bolt failure
			Bearing failure
			Joint failure
			Flange failure

			Flexible coupling failure
			Gasket failure
			Valve failure
		Hose failure	
		Lining failure	
		Pipeline failure	
		Seal failure	
		Tube failure	
		Equipment misalignment	
		Cooling tower collapse	
		Floating roof failure	
		Pump failure	
		Refractory failure	
		Shaft failure	
		Tank failure	
		Vessel failure	
	Safety Equipment failure	Alarm failure	
		Bursting Disc Failure	
		Safety Relief Valve Failure	
		Detection system failure	
	Ignition source	Fire (domino effect)	
		Chemical energy (domino effect)	
		Furnace, boiler, motor other unshielded hot surface	
		Heat transport canalisation	
		Magnifying glass	
		Resisting electric conductor (resistance, short-circuit)	

		Friction	
		Spark (mechanical)	
		Spark (static electricity)	
		Electric arc (defect of the electric installation, arc welding)	
		Impact	
		Combustion flame or heat	
		Natural event (lightning)	
		Hot missile (domino effect)	
	Utility failure	Air system failure	
		Fuel supply failure	Gas supply failure
		Hydraulic failure	
		Inert gas supply failure	Nitrogen supply failure
		Lubrication failure	
		Power supply failure	
		Steam failure	
		Water failure	
Organizational	Inadequate policy formulation		
	Inadequate policy implementation	Budget cuts	
		Communication inadequacy	
		Inadequate definition of line/staff responsibility	

	Risk assessment & Control System adequacy	Inadequate Hazard Analysis process	Inadequate risk analysis and tolerance criteria		
			Standards & codes requirement specifications		
			Design & development inadequacy	Inspection planning	
				Maintenance planning	
				Human factors (ergonomics) review	
				Emergency shutdown provision	
				Management of Change (MOC) provision	
				Operational readiness	Inadequate Training
					Pre Startup Safety Review (PSSR)
					Emergency provisions
					Lockouts
					Permit to Work System Inadequate
		Environment, Safety & Health (ES&H) program review			
Human Error	Operational error	Slip (Attentional failure)	Operator/Crew fatigue		
		Lapse (Memory failure)			
		Mistake (Lack of knowledge, faulty procedure)			
		Violation (Intention failure)	circumvention: 'does not happen to me' attitude		
			Drug misuse		

			Cigarette smoking		
		Operation inadequate, wrong, unnecessary	Inadequate cleaning		
			Inadequate testing		
			Inadequate identification		
			Inadequate installation		
			Inadequate inspection		
			Incorrect labelling		
		Operation Omitted	Atmosphere not tested		
			Bund Drain left open		
	Design error	Inadequate design			
		Inadequate standards applied	Machinery not guarded		
	Document error	Inadequate documentation	Map inadequate		
	Maintenance/repair	Inadequate isolation of equipment/system			
		Inadequate maintenance work operation			
	Procedure error	Faulty instructions			
		Incorrect cleaning procedure			
		Inadequate maintenance procedure			
		Safety procedures inadequate			
	Shift change	Inadequate communication			
	Tool unsatisfactory				
External Events	Deliberate acts	Arson			
		Terrorism			
		Vandalism			

		Sabotage	
		War	Bomb
			Missile
	External aggression	Plane Crash	
		Transportation accident	Road incidents
			Rail incidents
			Shipping incidents
			Plane crash
		Domino Effect	Fire (domino effect)
			Chemical energy (domino effect)
			Hot missile (domino effect)
		Excessive vibration	
		Friction heat	
	Natural Phenomenon	Weather effects	Rain
			Snow
			Ice
			Lightning
			Wind, gale, storm
		Avalanche	
		Earth movements	
		Flooding	Lifting of tanks
			fracture of pipelines

Consequences Taxonomy

People	Fatality			
	Injuries	Bruises/fractures		
		Asphyxiation		
		Electric Shock		
		Scalding/burning		
		Poisoning		
		Radioactive		
	Evacuation			
Environment	Polluting Spill	Ecological damage	Air	
			Water	
			Soil	
			Flora	
			Fauna	
Plant/Property/Equipment	Plant shutdown			
	Damage to equipment			
	Boilover			
	Capsize of platform			
	Leak/Rupture	Pipe leak/rupture		
		Tank leak/rupture		
	Collapse	Vessel collapse		
		Tankroof collapse		
	Blowout			
	Vessel overturned			
	Contamination of product			
	Product loss			
Fire	Liquid substance fire	Poolfire		
		Tank fire	Boilover and resulting poolfire	
		Conflagration		
		Toxic cloud		
		Ecological damage		
	Two-phase substance fire	Jetfire		

		Poolfire		
		Flashfire		
		Fireball (often after BLEVE)		
		Conflagration		
		Toxic cloud		
		Ecological damage		
	Gas/vapor substance fire	Flashfire		
		Conflagration		
		Toxic cloud		
		Ecological damage		
Implosion				
Explosion	Physical Explosion	Compressed gas/vapor explosion	Overpressure generation	
			Missiles ejection	
		BLEVE	Heat radiation	
			Overpressure generation	
			Missiles ejection	
		Rapid phase transition explosion	Overpressure generation	
	Chemical Explosion	Deflagration/Detonation	Vapor Cloud Explosion	Heat radiation
				Overpressure generation
			Aerosol explosion	Heat radiation
				Overpressure generation
			Gas explosion	Heat radiation
				Overpressure generation
				Missiles ejection
			Dust explosion	Heat radiation
				Overpressure generation
				Missiles ejection
			Condensed phase	Overpressure

			explosion	generation
				Missiles ejection
				Heat radiation
		Homogeneous chemical explosion	Runaway reaction and explosion	Overpressure generation
				Missiles ejection
				Toxic release
Internal fire/explosion				
Release of Substance	Toxic gas/vapor release			
	Flammable gas/vapor release	Gas/vapor substance fire		
		Explosion		
	Liquid/solid release to Air	polluting spill		
	Liquid/solid release to Water	polluting spill		
	Liquid/solid release to Ground	polluting spill		
Near miss				
Transport effect	Collision			
	Sinking			
	Derailment-consequence			
Other				

Barriers Taxonomy

Avoidance	Inherently *safe* design		
Prevention	Inherently *safer* Design		
	Mechanical Integrity		
	Preventive barriers	Inerting	
		Coating	
	Safety Management System	Management of Change	
		Procedures	
		Safety Culture	
		Periodic Inspection	
		Maintenance	
		Training	
	Basic Process Control System (BPCS)	Software driven steering of process	
	Operator Action	Supervision/Interven-tion on Alarms	
	Back-up system		
	Security System		
Protection	Facility spacing and layout		
	Detection system, alarms and operator intervention		
	Supervisor Intervention		
	Automatic SIS system	Detector, processor, actuator combination	explosion suppression
			isolation valve closing
	Isolation system	Fast acting valve	
		rotary valve	

		flame arrester	
	Physical Protection	Double Containment	
		Relief Valve	
	Ignition Control	Electrical Area Classification	
	Static electricity, lightning, and stray current protection		
Mitigation	Active Systems	Fire Suppression	Sprinkler
			Deluge system
			Foam system
			Fire extinguishers
		Explosion vents	
		Water curtains	
		Steam curtains	
	Passive Systems	Personal Protection Equipment	
		Dike/Bund	
		Containment and Drainage	
		Fireproofing/firewall	
		Sufficient spacing/distance to risk source	Facility spacing and layout
	Community/Plant Emergency Response	Decontamination	
		Fire Extinction	
	External fire & rescue service	Self-rescue possibilities	

		Access routes, water	
		Fire fighting	
		Medical treatment	

APPENDIX D

The modified incident taxonomy is shown below:

Chemicals Taxonomy

State	Gas		
	Liquid		
	Solid		
	Two-phase (gas-liquid)		
	Dust		
	Aerosol		
Hazard Classification	Physical Hazards	Explosivity	
		Reactivity	
		Water reactive	
		Flammability	
		Combustiblity	
		Oxidizing Agent	
		Pyrophoric	
		Pressure, High	Steam
		Vacuum	
		Corrosivity	
		Temperature, High	
		Temperature, Low (freezing)	
	Health Hazard	Acute Toxicity	
		Chronic Toxicity	

		Irritant/Sensitizer	
		Mutagenicity	
		Carcinogenicity	
		Asphyxiant	
	Environmental Hazards	Water	
		Air	
		Soil	
Type	Acids		
	Bases		
	Ammonia		
	Chlorine		
	Hydrocarbons	Crude Oil	
		Natural Gas	
		LPG	
		Gasoline, diesel, fuel oil or other petroleum product which is liq@ ambient conditions	
	Other organics		
	Inorganic compounds		
	Paints and Dyes		
	Polychlorinated Biphenyls		
	Polymers		
	Volatile Organic Compounds		
	Mixtures		
Use	Pesticides/Agricultural		
	Radiological	Radioactive Materials	
	Biological	Toxic Materials	

		Infectious substance	
	Medical	Magnetized Material	
		Hazardous Waste	
	Insecticide		
	Bactericide		
	Preservative		
	Solvent		
	Chemical reagent		
	Refrigerant		

Equipment Taxonomy

Storage Equipment	Tanks/Vessels	Atmospheric Tank	
		Pressurized Tank	
		Refrigerated Storage	
		Bullet Tank	
		Floating Roof Tank	
	Solids Storage	Hopper	
		Silo	
	Gas containers	Cylinder (compressed gas)	
	Other storage	Drum/barrel	
		Bag	
		Sampling container	
		Storage Depot/hangar	
		Salt Dome Storage	
		Underground storage	
		Pallets	
Process Equipment	Reactors	Batch	
		Continuous	
		Semi-batch	
		Liquid phase	
		Vapor phase	
		Catalytic	
		Stirred Tank	
		Tubular	
	Reforming Equipment	Catalytic Reformer	
		Other Reformer	

	Cracking Equipment	Catalytic Cracker	
		Cracking Furnace	
		Cracking Tower	
	Hydrotreaters	Hydrocracker	
		Fluid cracker	
	Heating and Cooling Equipment	Cooling Equipment	Air Cooler
			Cooler
			Cooling Tower
			Cooling water system
			Cryogenic equipment
			Refrigeration unit
			Fin fan cooler
		Heating Equipment	Concentrator
			Drier
			Reboiler
			Furnace
			Incinerator
			Kiln
			Oven
			Heating Coils
		Heat Exchanger	Shell and tube heat exchanger
			Plate heat exchanger
		Phase change equipment	Condenser
			Crystallizer
			Evaporator
			Vaporizer

	Relief Equipment & Systems	Pressure Relief Device	Pressure Safety Valve	
			Rupture Disc	
		Knockout Drum		
		Header System		
	Separation Equipment	Liquid-liquid separation	Decanting Vessel	
		Fractionating equipment	Distillation Column	Stripper
		Gas-liquid separation	Absorption	
			Scrubber	
			Stripper	
		Membrane separation		
		Solid-Gas Separation	Filter	
			Cyclone	
			Precipitator	
		Solid-liquid separation	Adsorption	
			Centrifuge	
			Dessicator	
			Filter	
			Ion exchange bed	
			Settling	
			Strainer	
		Solid-solid separation	Sieve	
	Pressure Raising or Reducing Equipment	Compressor	Strainer	
			Blower	
		Pump	Vacuum compressor	
			Canned pump	
		Pump or Compressor Components	Impellor	

			Pump bearing
			Pump drive
			Pump gearbox
			Pump gland
			Pump manifold
			Seal
		Ejector Equipment	
	Physical Processes Equipment	Mixer	
		Crystallization	
		Adsorption	
		Dryer	
	Packaging Equipment		
	Intermediate Equipment		
Pipes/Fittings/Valves	Piping	Process pipes	Body of Pipe
			Joint
			Elbow
			Tee
		Pipeline	Body of Pipe
			Joint
			Pipe seam
			Compressor station
			Regulator/metering system
			Scraper trap
			Bolted fitting
			Welded fitting
			Girth weld

			Sump
			Repair sleeve
			Service line
	Connector	Bellows	
		Bonnet Joint	
		Compression Fitting	
		Coupling	
		Expansion joint	
		Flange	
		Gasket	
	Valves	Manually operated valve	Control valve
			Drain cock
			Lever operated valve
		Automatic valve	Ball valve
			Bonnet valve
			Control valve
			Emergency Shutdown
			Foot valve
			Gate valve
			Non-return valve
			Petcock valve
			Steam trap
			Safety Relief valve
	Drains and Sewers	Drainage System	
		Effluent pond	
		Floating Roof Drain	
		Manhole	

		Sewer	
		Sump	
		Waste Water Treatment	
		Nozzle	
		Pipe Support	
	Hose		
	Insulation		
	Manifold		
Transfer Equipment	Pump		
	Compressor		
	Hoses		
	Gravity flow/chute		
	Mechanical transport	Conveyor	
		Crane	
	Transfer interface	Loading Equipment	
		Unloading Equipment	
Electrical Equipment	Transformer		
	Capacitor		
	Generator		
	Battery		
	Cable	Powerline	
	Electrical isolators and switches		
	Motor	Pump-motor	
		Stirrer-motor	
	Drilling Equipment		
	Junction box		

Process Control and Instrumentation	Computer		
	Control box		
	Control cable		
	Control room		
	Instrument or controller	Flow meter or control	
		Level meter or control	
		Process Analyzer	
		Pressure meter or control	
		Temperature meter or control	
		Viscosity meter	
		Weighing scale	
Ancillary Process Equipment	Hydraulic System		
	Disposal/Waste Treatment	Incinerator	
		Flare Stack	
		Waste water treatment	
Solids Processing Equipment	Ball mill		
	Chopper		
	Crumbling machine		
	Cutter		
	Extruder		
	Flaker		
	Granulator		
	Grater		
	Grinder		

	Pelletizer		
	Pebble mill		
Mechanical Handling Equipment	Conveyor		
	Chute		
	Fan		
	Hopper feeder		
	Lifting equipment	Crane	
		Elevator	
		Lifting Hoist	
		Winch	
		Wire Rope	
	Winding Machine		
Heating/Cooling for building			
Ventilation System			
Laboratory Equipment			
Power Plant	Engine		
	Turbine		
	Turbo Generator		
Vehicles & Mobile Equipment	Air transport	Aircraft	
		Helicopter	
	Loading/Unloading Point	Loading Arm	
		Ship or Shore Connection	
		Terminal	
	Marine Transport	Container Ship	
		Marine Freighter	
		Marine Tanker	

		Passenger Ship	
	Rails Transport	Freight Train	
		Rail Tanker	
		Rail Wagon	
	River Transport	River Barge	
		River Tanker	
	Road Transport	Bulldozer or JCB or Digger	
		Forklift Truck	
		Lorry	
		Road Tanker	
		Road Vehicle	
Facilities	Warehouse		
	Refinery		
	Production	Chemical Production	
		Food/Beverage	
		Pulp & paper	
		Metal	
	Tank Farm		
	Offshore Platform		
	Pumping Station		
	Water Purification	Wastewater	
		Feed water	
	Utility	Cogen	
		Boiler	
		Cryogenic Separation	

Operation Taxonomy

Construction		
Demolition		
Exploration		
Drilling	Offshore	
	Onshore	
Processing	Commissioning	
	Decommissioning	
	Startup	
	Shutdown	
	Emergency	
	Normal Operations	Draining
		Purging
		Sampling
		Valve Operation
		Mixing
	Materials Handling	
	Testing	
	Maintenance/Inspection	Cleaning
		Entry into Confined Space
		Hot Work
		Line Break
		Preparation for Maintenance
		Repair
	Batch Operation	
	Upset condition	

	Non-routine operation	
	Idle, Out-of-Service	
Transfer	Loading tank truck	
	Unloading tank truck	
	Material transfer	
Labwork		
Storage		
Warehousing		
Transportation	Rail	
	Road	
	Air	
	Marine	
	River	
	Pipeline	

Causes Taxonomy

Process Upsets	Overpressure	Internal overpressure (Gas material)	Combustion/explosion causes overpressure
			External HP source connected causes overpressure
			Internal high pressure causes overpressure
			Overcompression causes overpressure
			Runaway (side) reaction causes overpressure
		Internal overpressure (liquid)	Overfilling vessel causes overpressure
			Pump causes overpressure
			Runaway (side) reaction causes overpressure
			Thermal expansion of liquid filled vessel causes overpressure
			Back pressure wave causes overpressure
			Combustion/explosion causes overpressure
			Roll-over of vessel contents causes overpressure
			Freezing and expansion of aqueous phase
		Internal overpressure (solid material)	Dust explosion overpressures enclosure
			Unstable solid explodes
	Underpressure	Underpressure (pressure below the containment limit of the vessel)	Chemical reaction consuming gases

			Condensation
			Decrease of pressure due to a decrease of the temperature
			Fast emptying of the vessel
		Underpressure (pressure below the containment limit of the roof)	Condensation
			Decrease of pressure due to a decrease of the temperature
			Chemical action consuming gases
			Fast emptying of the vessel or tank
	Overheating	High temperature	Fire (domino effect)
			Chemical energy (domino effect)
			Furnace, boiler, motor other unshielded hot surface
			Heat transport canalisation
			Resisting electric conductor (resistance, short-circuit)
			Electric arc (defect of the electric installation, arc welding)
			Combustion flame or heat (gas flame welding, smoking)
			Natural event (lightning)
			Hot missile (domino effect)
		Temporary (or permanent) energy source	Fire (domino effect)
			Chemical energy (domino effect)
			Furnace, boiler, motor other unshielded hot surface
			Heat transport canalisation
			Magnifying glass

			Resisting electric conductor (resistance, short-circuit)
			Friction (conveyors, doors, mechanical devices)
			Spark (mechanical)
			Spark (static electricity)
			Electric arc (defect of the electric installation, arc welding)
			Impact
			Combustion flame or heat (gas flame welding, smoking)
			Natural event (lightning)
			Hot missile (domino effect)
	Overfilling	Overflow	
	Underfilling		
	Incorrect flow	Incorrect flow rate	Flow rate too high
			Flow rate too low
		Backflow	
		Overflow	
		Reverse flow	
	Incorrect temperature	High temperature	
		Low temperature	
	Exothermic reaction		
	Flameout		
	Frothing		
	Leak	Air leaking into system	
		Flange leak	
		Gasket leak	
		Joint leak	

		Relief valve Leak	
	Low level of catalyst		
	Roll-over stratified liquid layers (due to temperature/density differences)	Vapor cloud release	
Chemical Causes	Accidental mixing/Contamination	Contact with a combustible (reducing) substance	Presence of an incompatible reagent in the storage room
			Leak, entrainment from a nearby pipe or vessel (domino effect)
			Accidental contamination
			The reducing substances is part of the stored/used chemical (organic peroxide)
		Contact with an incompatible reagent	Presence of an incompatible reagent in the storage room
			Leak, entrainment from a nearby pipe or vessel (domino effect)
			Contamination by natural elements of the environment (air, water)
			Accidental contamination
		Presence of a catalyst (if involved)	Abnormal presence of a catalyst
			Leak, entrainment from a nearby pipe or vessel (domino effect)
			Accidental contamination
		Presence of air vector	Aeration
			Wind
			Gas escaping from a breach or leak (domino effect)

			Air projection due to fall, collapse or explosion of a nearby equipment (domino effect)
			Exhaust gases
		Presence of liquid vector	Liquid escaping from a leak or breach
			Extinction water
			Natural event
			Other water sources
	Oxygen Enrichment	Presence of oxidizing substance	Normal storage/use
			Abnormal storage/use
			Leak, entrainment from a nearby equipment (domino effect)
			Other accidental contamination
	Potentially mobile material	Fragmented material	
		Soluble material	
		Material susceptible to emit vapours (e.g. solvent)	
	Hazardous material storage	Storage of a flammable product	
		Storage of a reaction sensitive chemical	
		Storage of a thermal decomposition sensitive chemical	
	Hazardous material use	Use of a highly reactive substance	
		Use of an explosive material	
		Use of a flammable substance	
		Use of a toxic substance	
	Chemical missing	Low level of catalyst	

		Lack of stabilizer/inhibitor		
		Oxygen deficiency		
	Incompatible material of construction			
	Incorrect chemical concentration			
	Unwanted chemical reaction	Runaway reaction		
		Decomposition		
		Autoignition		
		Spontaneous combustion	Hypergolic liquids mixing	
		Thermite reaction	Al + Fe2O3 reaction	
		Unwanted polymerization		
Equipment Failure	Material of construction failure	Brittle rupture	Brittle structure	
			Impact	
		Degradation of the mechanical properties leads to incapacity to sustain high pressure	Corrosion	
			Erosion	
			Fatigue	
			Fracture	Crack
				Creep
			Thermal weakening (lowering of the tensile or compressive stength under the effect of high temperature)	
			Other chemical action of the environment (solvent on plastic structure)	

			Other physical action on the environment (light on plastics)
			Aging
			Rusting
			Stress
			Embrittlement
		Insufficient Initial mechanical properties of the structure (below normal use conditions)	Inappropriate/insufficient material
			Inappropriate dimnensions
			Inappropriate assembling
		Excessive external stress	Overloading
			High amplitude vibrations
			Dilation
			External overpressure
			Shear stress
	Instrumentation failure		
	Control failure	Computer system failure	
	Equipment missing	Open manhole	
	Incorrect equipment installed		
	Additional equipment installed	insulating plate	
	Electrical Equipment Failure	Power supply failure	
		Generator failure	
		Motor failure	
		Short circuit	
		Lack of earthing	

		Spark	
		Electric arc (defect of the electric installation, arc welding)	
	Mechanical equipment failure	Accidental opening	Disconnect during filling
			Valve opened by mistake
			Valve left open by mistake
			External force causes accidental opening
			Valve blocked
			Seal, joint loss of effectiveness
			Safety valve, safety relief device
			Fail to clear out contents before opening containment
			Wrong part (containing hazardous material) worked on
			Operation started when containment open
			Uncontrolled flow during sampling/draining
		Leak	Air leaking into system
			Flange leak
			Gasket leak
			Joint leak
			Relief valve Leak
		Blower failure	
		Fittings failure	Bolt failure
			Bearing failure
			Joint failure
			Flange failure

			Flexible coupling failure
			Gasket failure
			Valve failure
		Hose failure	
		Lining failure	
		Pipeline failure	
		Seal failure	
		Tube failure	
		Equipment misalignment	
		Cooling tower collapse	
		Floating roof failure	
		Pump failure	
		Refractory failure	
		Shaft failure	
		Tank failure	
		Vessel failure	
	Safety Equipment failure	Alarm failure	
		Bursting Disc Failure	
		Safety Relief Valve Failure	
		Detection system failure	
	Ignition source	Fire (domino effect)	
		Chemical energy (domino effect)	
		Furnace, boiler, motor other unshielded hot surface	
		Heat transport canalisation	
		Magnifying glass	
		Resisting electric conductor (resistance, short-circuit)	

		Friction	
		Spark (mechanical)	
		Spark (static electricity)	
		Electric arc (defect of the electric installation, arc welding)	
		Impact	
		Combustion flame or heat	
		Natural event (lightning)	
		Hot missile (domino effect)	
	Utility failure	Air system failure	
		Fuel supply failure	Gas supply failure
		Hydraulic failure	
		Inert gas supply failure	Nitrogen supply failure
		Lubrication failure	
		Power supply failure	
		Steam failure	
		Water failure	
Organizational	Inadequate policy formulation		
	Inadequate policy implementation	Budget cuts	
		Communication inadequacy	
		Inadequate definition of line/staff responsibility	
	Inadequate Process Safety Management	Inadequate PHA	
		Standards & codes requirement not followed	
		Process Design review	

		Inspection\Maintenance planning	
		Human factors (ergonomics) review	
		Emergency shutdown provision	
		Management of Change (MOC) provision	
		Operational readiness	Training & Performance Checks
			Pre Startup Safety Review (PSSR)
			Emergency provisions
			Lockouts
			Permit to Work System Inadequate
	Audits & Corrective Actions		
Human Error	Operational error	Slip (Attentional failure)	Operator/Crew fatigue
		Lapse (Memory failure)	
		Mistake (Lack of knowledge, faulty procedure)	
		Violation (Intention failure)	circumvention: 'does not happen to me' attitude
			Drug misuse
			Cigarette smoking
		Operation inadequate, wrong, unnecessary	Inadequate cleaning
			Inadequate testing
			Inadequate identification
			Inadequate installation
			Inadequate inspection
			Incorrect labelling
		Operation Omitted	Atmosphere not tested

			Bund Drain left open
	Design error	Inadequate design	
		Inadequate standards applied	Machinery not guarded
	Document error	Inadequate documentation	Map inadequate
	Maintenance/repair	Inadequate isolation of equipment/system	
		Inadequate maintenance work operation	
		Damage caused by maintenance work	
	Procedure error	Faulty instructions	
		Incorrect cleaning procedure	
		Inadequate maintenance procedure	
		Safety procedures inadequate	
	Shift change	Inadequate communication	
	Tool unsatisfactory		
External Events	Deliberate acts	Arson	
		Terrorism	
		Vandalism	
		Sabotage	
		War	Bomb
			Missile
	External aggression	Plane Crash	
		Transportation accident	Road incidents
			Rail incidents
			Shipping incidents
			Plane crash

		Domino Effect	Fire (domino effect)
			Chemical energy (domino effect)
			Hot missile (domino effect)
		Excessive vibration	
		Friction heat	
	Natural Phenomenon	Weather effects	Rain
			Snow
			Ice
			Lightning
			Wind, gale, storm
		Avalanche	
		Earth movements	
		Flooding	Lifting of tanks
			fracture of pipelines
Other			
Initiating Event	Design Related		
	Operations Related		
	Maintenance		
	Other Factors		
	Metal Fatigue		
	Corrosion		
	Embrittlement		
	Vibration		

Consequences Taxonomy

People	Fatality		
	Injuries	Bruises/fractures	
		Asphyxiation	
		Electric Shock	
		Scalding/burning	
		Poisoning	
		Radioactive	
	Evacuation		
	At risk		
Environment	Ecological damage	Air	
		Water	
		Soil	
		Flora	
		Fauna	
Plant/Property/Equipment	Plant shutdown		
	Damage to equipment/installation		
	Boilover		
	Capsize of platform		
	Leak/Rupture	Pipe leak/rupture	
		Tank leak/rupture	
	Collapse	Vessel collapse	
		Tankroof collapse	
	Blowout		
	Vessel overturned		
	Contamination of product		

	Product loss		
Fire	Liquid substance fire	Poolfire	
		Tank fire	Boilover and resulting poolfire
		Conflagration	
		Toxic cloud	
		Ecological damage	
	Two-phase substance fire	Jetfire	
		Poolfire	
		Flashfire	
		Fireball (often after BLEVE)	
		Conflagration	
		Toxic cloud	
		Ecological damage	
	Gas/vapor substance fire	Flashfire	
		Conflagration	
		Toxic cloud	
		Ecological damage	
Implosion			
Explosion	Physical Explosion	Compressed gas/vapor explosion	Overpressure generation
			Missiles ejection
		BLEVE	Heat radiation
			Overpressure generation
			Missiles ejection
		Rapid phase transition explosion	Overpressure generation
	Chemical Explosion	Vapor Cloud Explosion	Heat radiation

			Overpressure generation
		Aerosol explosion	Heat radiation
			Overpressure generation
		Gas explosion	Heat radiation
			Overpressure generation
			Missiles ejection
		Dust explosion	Heat radiation
			Overpressure generation
			Missiles ejection
		Condensed phase explosion	Overpressure generation
			Missiles ejection
			Heat radiation
		Runaway reaction and explosion	Overpressure generation
			Missiles ejection
			Toxic release
Internal fire/explosion			
Release of Substance	Toxic gas/vapor release	Contamination	
	Flammable gas/vapor release	Gas/vapor substance fire	
		Explosion	
	Liquid/solid release to Air	polluting spill	
	Liquid/solid release to Water	polluting spill	
	Liquid/solid release to Ground	polluting spill	
Near miss			
Transport effect	Collision		

	Sinking		
	Derailment-consequence		
Other			

Layers of Protection (LOP) Taxonomy

Prevention	Inherently safer Design		
	Mechanical Integrity		
	Preventive barriers	Inerting	
		Coating	
	Safety Management System	Management of Change	
		Procedures	
		Safety Culture	
		Periodic Inspection	
		Maintenance	
		Training	
	Basic Process Control System (BPCS)	Software driven steering of process	
	Operator Action	Supervision/Interven-tion on Alarms	
	Back-up system		
	Security System		
Protection	Facility spacing and layout		
	Detection system, alarms and operator intervention		
	Supervisor Intervention		
	Automatic SIS system	Detector, processor, actuator combination	explosion suppression
			isolation valve closing
	Isolation system	Fast acting valve	

		rotary valve	
		flame arrester	
	Physical Protection	Double Containment	
		Relief Valve	
	Ignition Control	Electrical Area Classification	
	Static electricity, lightning, and stray current protection		
Mitigation	Active Systems	Fire Suppression	Sprinkler
			Deluge system
			Foam system
			Fire extinguishers
		Explosion vents	
		Water curtains	
		Steam curtains	
	Passive Systems	Personal Protection Equipment	
		Dike/Bund/catchment	
		Containment and Drainage	
		Fireproofing/firewall	
		Sufficient spacing/distance to risk source	Facility spacing and layout
	Community/Plant Emergency Response	Decontamination	
		Fire Extinction	

		Emergency shutdown	
		Rescue possibilities	
		Access routes, water	
		Gas detection/ dispersion prediction	
	External fire & rescue service	Fire fighting	
		Medical treatment	
		Repair of equipment	
		Evacuation	

Recommendations & Actions Taxonomy

Changes to design	Improve protection of equipment/piping/valve
	Software-driven steering of process
	Adaptation of material of construction for equipment to substance properties
	Inherently safer design
Changes to maintenance procedures	
Changes to operating procedures	
Better observation of procedures	
Better training of operators/maintenance workers/supervisors	
Changes to raw materials/solvents/catalysts	
Changes to codes/standards	
Changes to company culture	
Better safety management	
Installation shutdown	
Under investigation	

APPENDIX E

The following are the codes used for developing the search system:

Code for the word search:

```
public class HelloLucene {
  private Directory index;
  private IndexWriter w;
  private static boolean flag = false;

  //The function which does the indexing of the text files
  public void doIndexing() {
    try {
      index = new RAMDirectory();
      w = new IndexWriter(index, new StandardAnalyzer(), true, new MaxFieldLength(25000));
      File directory = new File("D:\\work\\txt");
      if (directory.isDirectory()) {
        File[] filesInDirectory = directory.listFiles();
        for (int i = 0; i < filesInDirectory.length; i++) {
          File file = filesInDirectory[i];
          if (file.getName().endsWith(".txt")) {
            String contents = "";
            BufferedReader reader = new BufferedReader(new FileReader(file));
            String line = reader.readLine();
            while (line != null) {
              contents += line;
              line = reader.readLine();
            }
            addDoc(w, contents, file.getName());
          }
        }
      }
      w.close();
    }
    catch (Exception e) {
      e.printStackTrace();
    }
  }

  //The function which does the actual Searching.
  public String doSearch(String queryString) throws CorruptIndexException,
                LockObtainFailedException, IOException,
```

```
                     ParseException {
    if (!flag) {
      doIndexing();
      flag = true;
    }
    queryString = "\"" + queryString + "\"";
    Query q = new QueryParser("title", new StandardAnalyzer()).parse(queryString);
    IndexSearcher s = new IndexSearcher(index);
    TopDocs docs = s.search(q, null, 800);
    ScoreDoc [] hits = docs.scoreDocs;
    String returnString = "";
    for(ScoreDoc scoreDoc : hits) {
      String fileName = s.doc(scoreDoc.doc).get("name");
      returnString += fileName.substring(0, fileName.indexOf(".")) + "@@@";
    }
    s.close();
    return returnString;
  }

  private void addDoc(IndexWriter w, String value, String name) throws IOException {
    Document doc = new Document();
    doc.add(new Field("title", value, Field.Store.YES, Field.Index.ANALYZED));
    doc.add(new Field("name", name, Field.Store.YES, Field.Index.ANALYZED));
    w.addDocument(doc);
  }
}
```

Program to create the folder search tree:

```
public class WebServiceTest {
  private static String excelContents = "";

       //Reads an Excel file to create a Folder tree.
  public String readExcelFile() {
    if (excelContents.length() > 10) {
      return excelContents;
    }
    String fileName = "D:\\1.xls";
    File file = new File(fileName);
    Workbook workbook;
    Sheet sheet;
    String returnValue = "";
    String divider = "";
```

```
try {
   workbook = Workbook.getWorkbook(file);
   int numSheets = workbook.getNumberOfSheets();
   for (int k = 0; k < numSheets; k++) {
      String sheetNumber = "#LEVELZERO#";
      sheet = workbook.getSheet(k);
      returnValue += sheetNumber + sheet.getName();
      for (int i = 0; i < sheet.getRows(); i++) {
         for (int j = 0; j < sheet.getColumns(); j++){
            divider = "#LEVEL";
            switch (j) {
            case 0:
               divider += "ONE#";
               break;
            case 1:
               divider += "TWO#";
               break;
            case 2:
               divider += "THREE#";
               break;
            case 3:
               divider += "FOUR#";
               break;
            case 4:
               divider += "FIVE#";
               break;
            case 5:
               divider += "SIX#";
               break;
            case 6:
               divider += "SEVEN#";
               break;
            }
            if (sheet.getCell(j, i).getContents() != null
               && sheet.getCell(j, i).getContents().length() > 0){
                  returnValue += divider + sheet.getCell(j, i).getContents();
               }
         }
      }
   }
} catch (IOException e) {
   e.printStackTrace();
} catch (BiffException e) {
   e.printStackTrace();
}
```

```
        excelContents = returnValue;
        return returnValue;
    }
}
```

Code for word search within a folder:

```
public class InFolderWordSearch {
    private String[] getFiles(String fileNamesList) {
        String[] returnList = fileNamesList.split(":")[1].split(",");
        for (int i = 0; i < returnList.length; i++) {
            returnList[i] = returnList[i].trim();
        }
        return returnList;
    }

    private Directory index;
    private IndexWriter w;
            //Function to index the required files
    private void indexFiles(String[] fileNamesList) {
        try {
            index = new RAMDirectory();
            w = new IndexWriter(index, new StandardAnalyzer(), true, new MaxFieldLength(25000));
            for (int i = 0; i < fileNamesList.length; i++) {
                File file = new File("D:\\work\\txt\\" + fileNamesList[i] + ".txt");
                String contents = "";
                BufferedReader reader = new BufferedReader(new FileReader(file));
                String line = reader.readLine();
                while (line != null) {
                    contents += line;
                    line = reader.readLine();
                }
                addDoc(w, contents, file.getName());
            }
            w.close();
        }
        catch (Exception e) {
            e.printStackTrace();
        }
    }

    private void addDoc(IndexWriter w, String value, String name) throws IOException {
```

```
    Document doc = new Document();
    doc.add(new Field("title", value, Field.Store.YES, Field.Index.ANALYZED));
    doc.add(new Field("name", name, Field.Store.YES, Field.Index.ANALYZED));
    w.addDocument(doc);
  }

       //Main function - uses the index to do a query and return the results
  public String returnInFolderSearchResults(String queryString, String selectedItem) {
    queryString = "\"" + queryString + "\"";
    String returnString = "";
    try {
      File file = new File("D:\\8.xls");
      Workbook wb = Workbook.getWorkbook(file);
      Sheet sheet = wb.getSheet(0);
      int numRows = sheet.getRows();
      for (int i = 0; i < numRows; i++) {
        String data = sheet.getCell(0, i).getContents();
        if (data.equalsIgnoreCase(selectedItem)) {
          String fileNamesList = sheet.getCell(0, i).getCellFeatures().getComment();
          String[] fileNames = getFiles(fileNamesList);
          indexFiles(fileNames);
          Query q = new QueryParser("title", new StandardAnalyzer()).parse(queryString);
          IndexSearcher s = new IndexSearcher(index);
          TopDocs docs = s.search(q, null, 100);
          ScoreDoc [] hits = docs.scoreDocs;
          for(ScoreDoc scoreDoc : hits) {
            returnString += fileName.substring(0, fileName.indexOf(".")) + "@@@";
          }
          s.close();
          break;
        }
      }
    }
    catch (Exception e) {
      e.printStackTrace();
    }
    return returnString;
  }
}
```

Code to read MARS pages and save them as html and text files:

```
public class ReadAllWebPages {
```

```
        //Function to save all the MARS database files as TEXT files into the hard disk (these
files are used for indexing purposes)
    public void ReadAllTXT() {
        File file = new File("D:\\3.xls");
        try {
            Workbook wb = Workbook.getWorkbook(file);
            Sheet sheet = wb.getSheet(0);
            int rows = sheet.getRows();
            String data;
            String writeFileName;
            for (int i = 1; i < rows; i++) {
                writeFileName = "D:\\work\\txt\\";
                data = sheet.getCell(0, i).getContents();
                writeFileName += data + ".txt";
                returnFileContents(data, writeFileName);
            }
        }
        catch (Exception e) {
            e.printStackTrace();
        }
    }

        //Function to save all the MARS database files onto the hard disk.
    public void readAllHTML() {
        File file = new File("D:\\3.xls");
        try {
            Workbook wb = Workbook.getWorkbook(file);
            Sheet sheet = wb.getSheet(0);
            int rows = sheet.getRows();
            String data;
            String writeFileName;
            for (int i = 1; i < rows; i++) {
                writeFileName = "D:\\work\\html\\";
                data = sheet.getCell(0, i).getContents();
                writeFileName += data + ".html";
                returnFileContents(data, writeFileName);
            }
        }
        catch (Exception e) {
            e.printStackTrace();
        }
    }

    public String returnFileContents(String sNum, String writeFileName) {
        String urlString
```

```
        = "http://mahb-
srv.jrc.it/typo3/fileadmin/eMARS_Site/PhpPages/ViewAccident/ViewAccidentPublic.php?accid
ent_code=" + sNum;
      try {
        URL url = new URL(urlString);
        BufferedReader reader = new BufferedReader(new
InputStreamReader(url.openStream()));
        String line, page = "";
        line = reader.readLine();
        while (line != null) {
          page += line + " ";
          line = reader.readLine();
        }
        reader.close();
        File writeFile;
        writeFile = new File(writeFileName);
        BufferedWriter writer =
          new BufferedWriter(new OutputStreamWriter(new FileOutputStream(writeFile)));
        writer.write(page);
        writer.close();
      }
      catch (Exception e) {
        e.printStackTrace();
      }
      return null;
    }

    private String output = "";
         //Recursive Function, which goes through all the nodes and finds only the needed ones.
         public int findChildrenRecursively(Node node, int level) {
        NodeList nodeList = node.getChildren();
        String temp;
        if (node.getChildren() == null) {
          String text = node.getText().trim();
          text = text.replaceAll(" ", " ");
          text = text.trim();
          if (text.endsWith("-->") && text.startsWith("<!--"))
            return level;
          if (text.startsWith("<!--")) {
            text = text.substring(text.lastIndexOf("EN-GB") + 7);
          }
          if (text.contains("clearpixel.gif"))
            return level;
          if (text.equalsIgnoreCase("b"))
            return level;
```

```
    if (text.startsWith("<div"))
      return level;
    if (text.startsWith("td id=") || text.startsWith("td class="))
      return level;
    if (text.equalsIgnoreCase("br"))
      return level;
    if (text.equalsIgnoreCase("strong"))
      return level;
    if (text.equalsIgnoreCase("td"))
      return level;
    if (text.equalsIgnoreCase("</div>"))
      return level;
    if (text.contains("!DOCTYPE"))
      return level;
    if (text.startsWith("link rel"))
      return level;
    if (text.equalsIgnoreCase("title"))
      return level;
    if (text.equalsIgnoreCase(" "))
      return level;
    if (text.startsWith("[if "))
      return level;
    if (text.equalsIgnoreCase("em"))
      return level;
    if (text.equalsIgnoreCase("body"))
      return level;
    if (text.equalsIgnoreCase("[endif]"))
      return level;
    if (text.startsWith("table border="))
      return level;
    if (text.startsWith("<style "))
      return level;
    if (text.startsWith("BODY {"))
      return level;
    if (text.startsWith("img src="))
      return level;
    output += text + " ";
    return level;
  }
  for (int i = 0; i < nodeList.size(); i++) {
    Node tempNode = nodeList.elementAt(i);
    temp = tempNode.getText().trim();
    if (temp.startsWith("meta"))
      continue;
    if (!temp.equals("") && !(temp.charAt(0) == '/')) {
```

```
                level += 1;
                findChildrenRecursively(tempNode, level);
                level -= 1;
            }
        }
        return level;
    }
}
```

APPENDIX F

The user interface (UI) screen shots for the database search system are shown here.

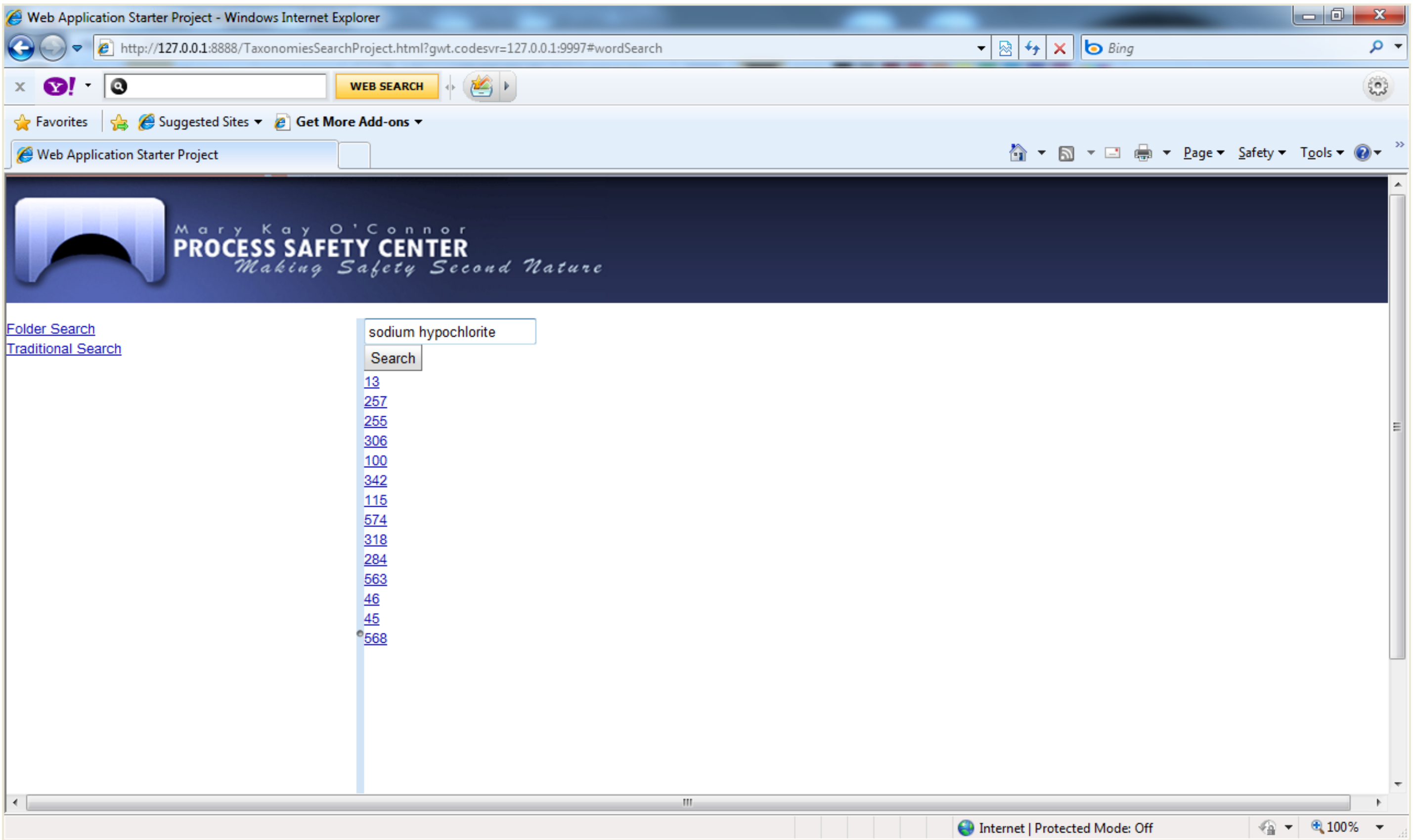

Figure F. 1 UI for the word search showing search results for sodium hypochlorite.

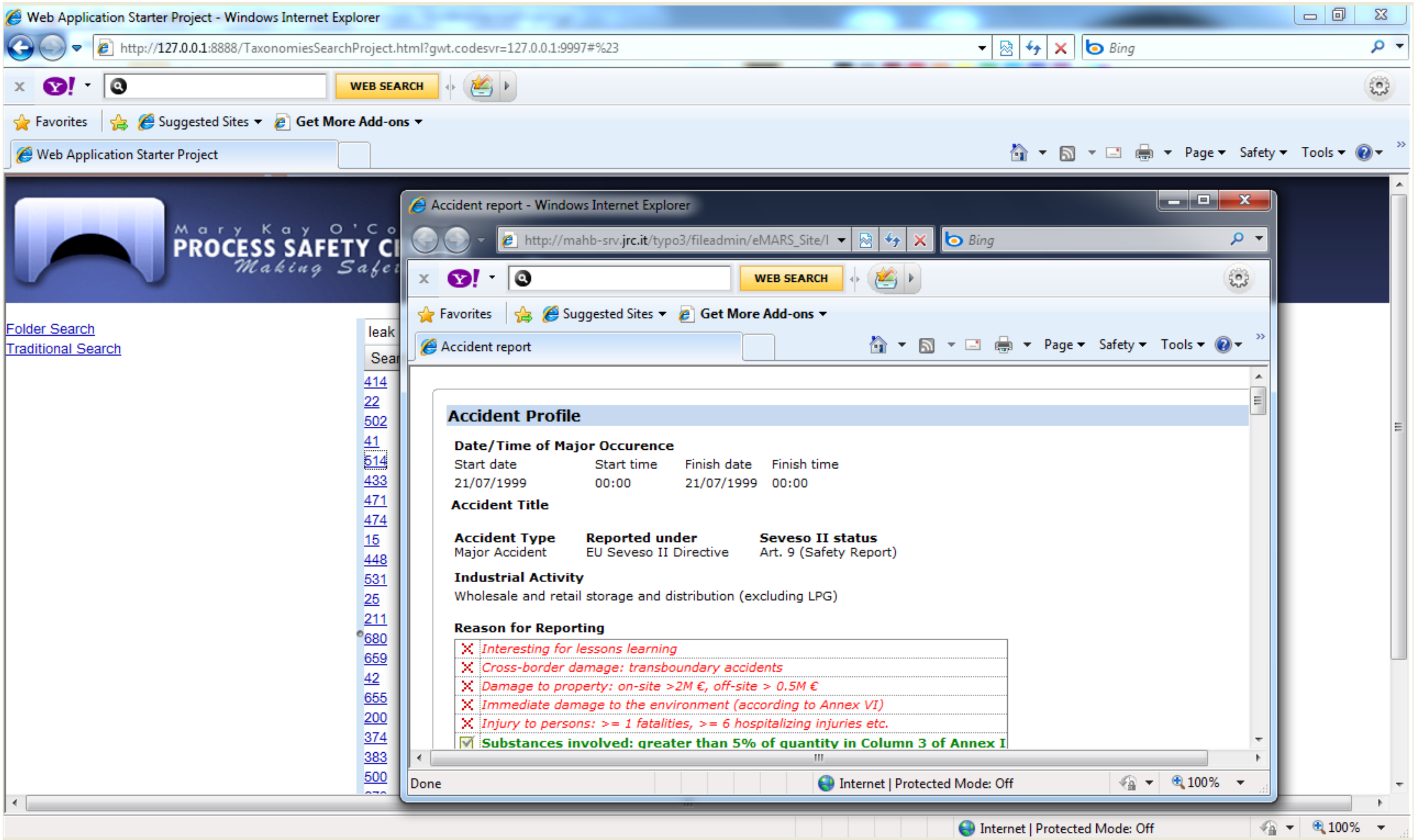

Figure F. 2 Screenshot of pop up window of incident information after clicking on one of the word search results.

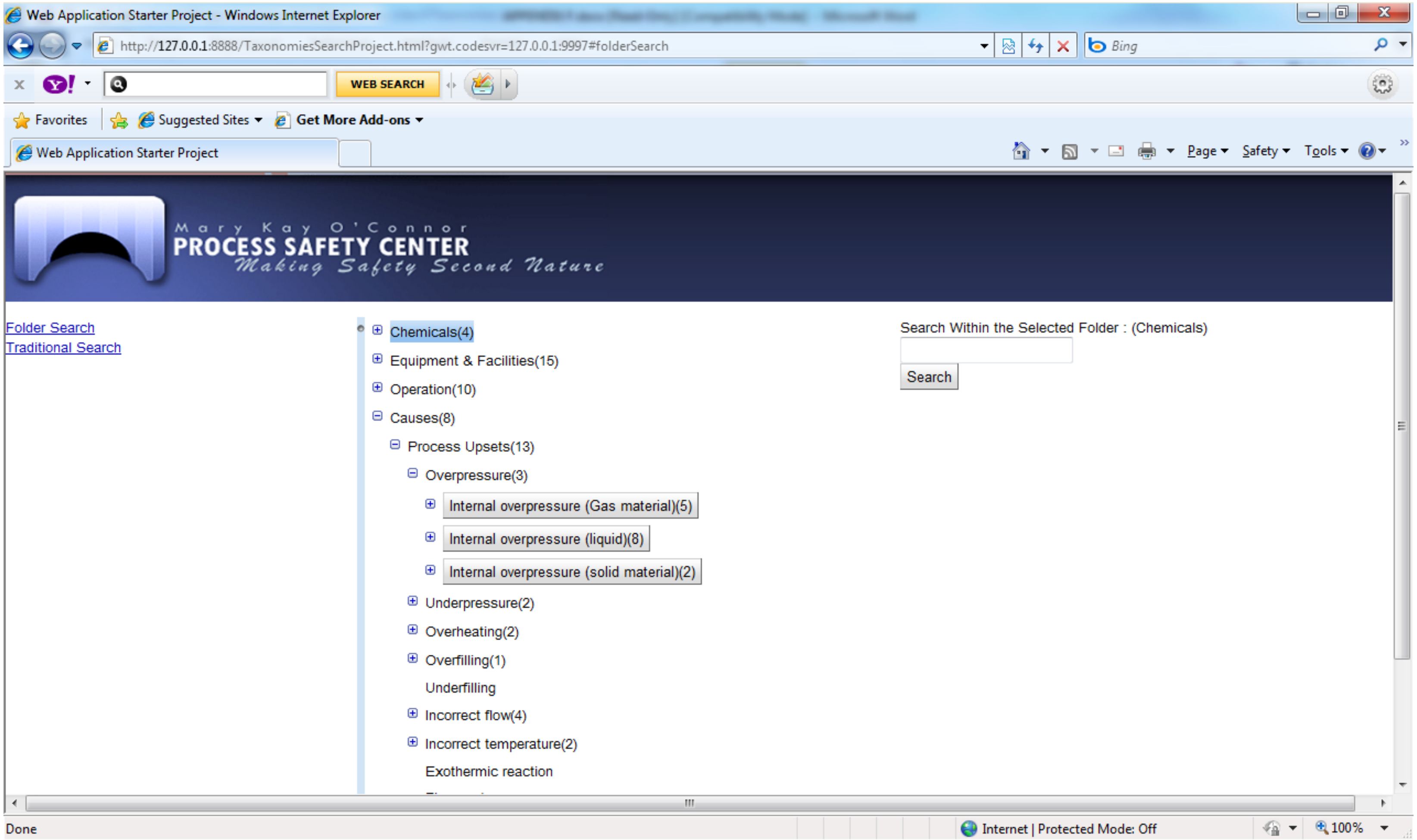

Figure F. 3. UI of clickable folders in the folder search system.

In Figure F.3, the ‘plus’ tab left of the node is for unfolding the node. The node that looks like a button are linked to ARAMIS fault tree screens. The number in parenthesis on the right hand side of each node is the number of items to the next level.

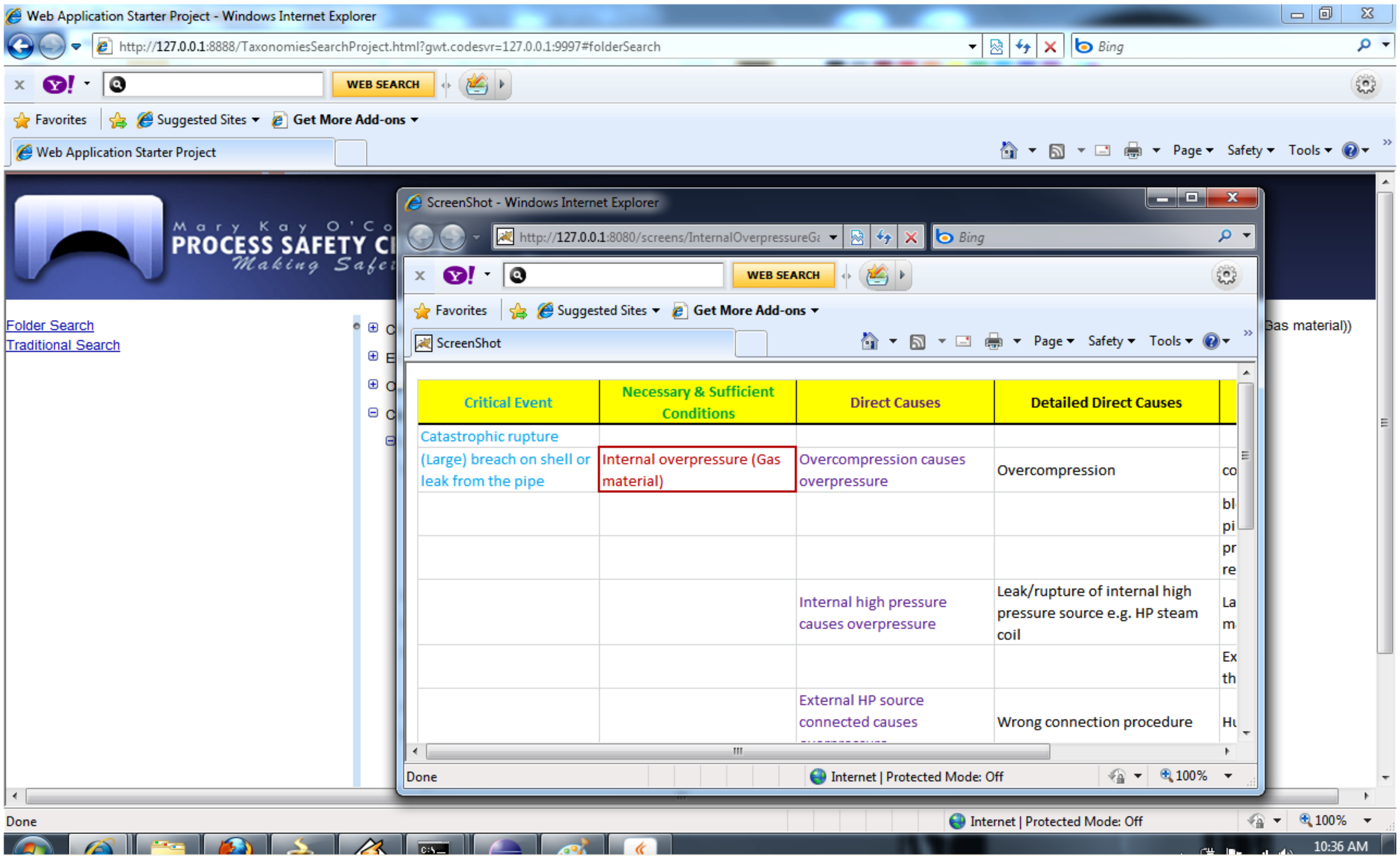

Figure F. 4 Screen shot of pop up window for clickable ARAMIS fault tree linked to causes taxonomy.

Figure F.4 shows the screen shot of part of the ARAMIS fault tree pop up window, when the user clicked on the button of *Internal overpressure (Gas material)* node.

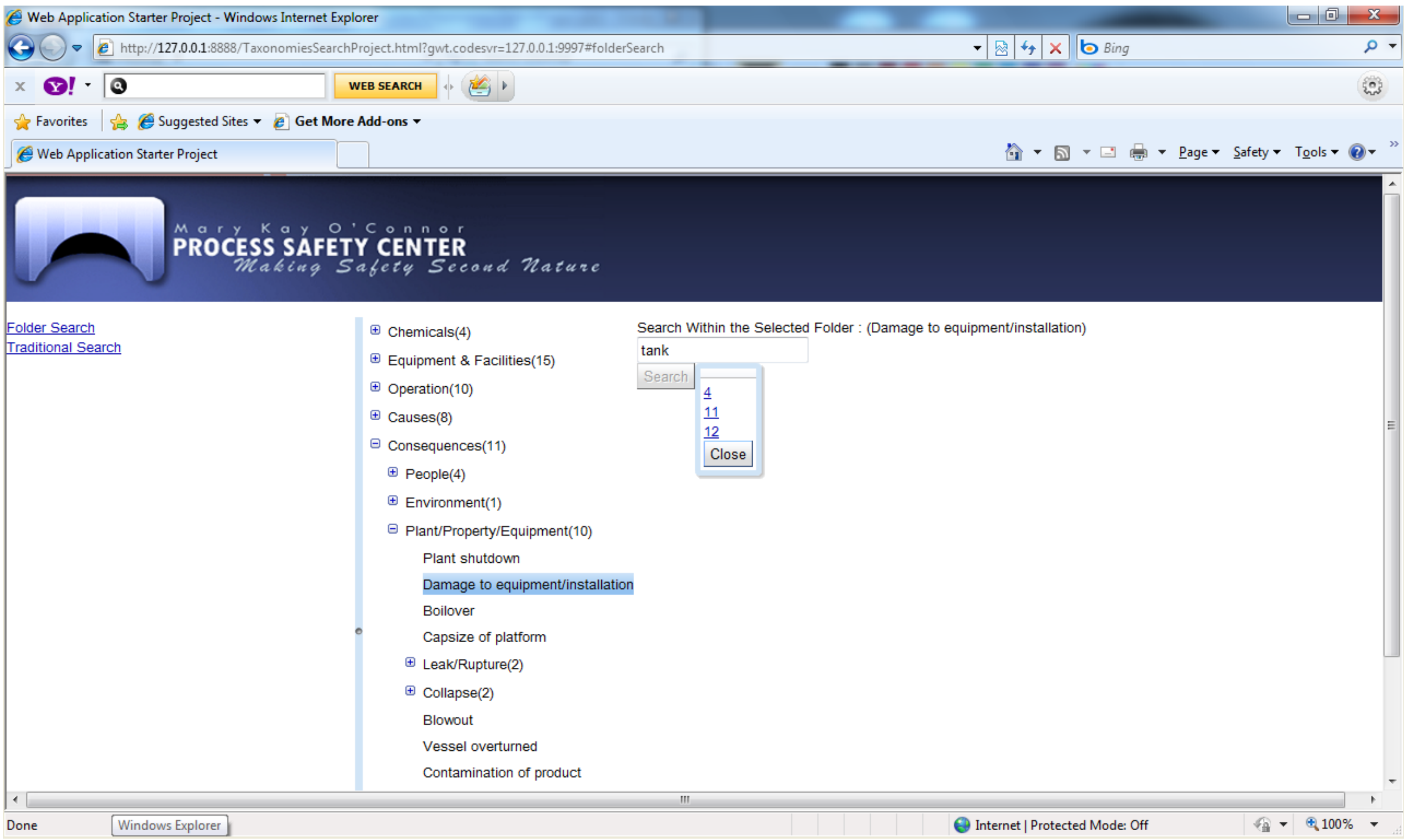

Figure F. 5 Screen shot of results of a word search within a folder.

In Figure F.5, the links in the little pop up window are incident files that are clickable to pop up incident information.

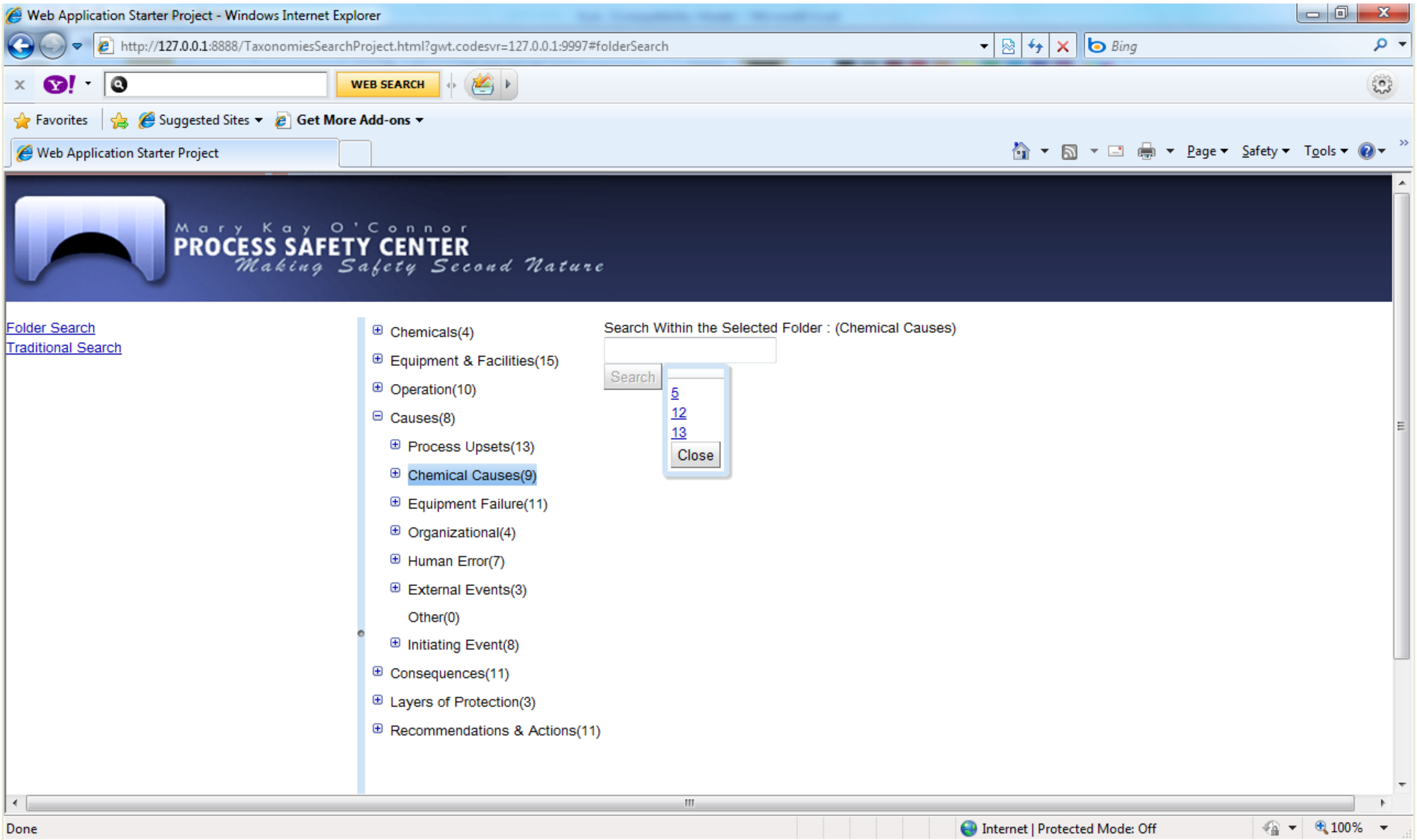

Figure F. 6 Screenshot of a blank word search on a folder to retrieve all records within that folder (Chemical Causes)

VITA

Sara Shammi Khan received her Bachelor of Science degree in Radiological Health Engineering from Texas A&M University at College Station in 2006. In August 2007, she joined the Mary Kay O'Connor Process Safety Center at Texas A&M University to pursue a Master of Science degree in Safety Engineering, under the guidance of Dr. M. Sam Mannan, and received her degree in 2010.

Sara Khan's permanent address is:
House #33, Road-6, Sector-4
Uttara Model Town
Dhaka-1230
Bangladesh.