33rd Annual INCOSE international symposium
hybrid event
Honolulu, HI, USA
July 15 - 20, 2023

MIKA: Manager for Intelligent Knowledge Access Toolkit for Engineering Knowledge Discovery and Information Retrieval

Sequoia Andrade
HX5 - NASA Ames Research Center,
Mail Stop 269-1
Moffett Field, CA, 94035-001
(650)604-3358
Sequoia.r.Andrade@nasa.gov

Hannah Walsh, PhD
NASA Ames Research Center
Mail Stop 269-1
Moffett Field, CA 94035-0001
(650)604-4532
Hannah.s.Walsh@nasa.gov

Abstract. Repositories of safety reports are often underutilized and only analyzed manually by trained experts, despite safety management systems requiring reports. These collections of documents contain a wealth of information from past projects and operations that could improve system safety and design. Advances in natural language processing techniques have improved information extraction and retrieval in consumer technology, biomedicine, and finance, for instance, but have not been applied to engineering documents on the same scale. To this end, the Manager for Intelligent Knowledge Access (MIKA) open-source toolkit has been developed for rapid knowledge discovery and information retrieval in safety engineering applications. The MIKA toolkit uses state-of-the-art natural language processing algorithms and allows a user to apply these methods to their own dataset. This paper describes the MIKA toolkit and its two primary capabilities, knowledge discovery and information retrieval, and demonstrates the toolkit via a case study on National Transportation Safety Board (NTSB) reports.

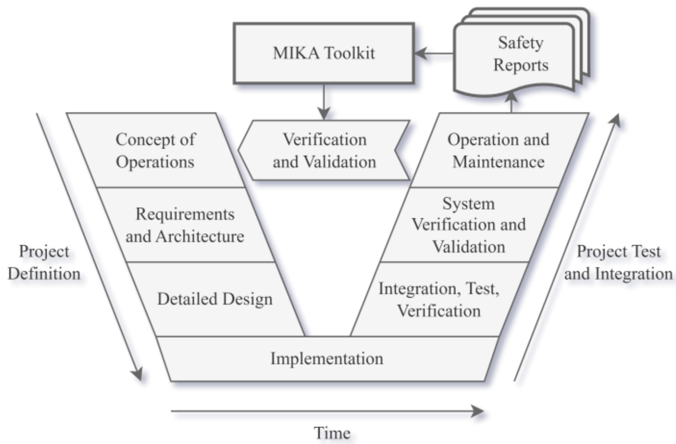
Introduction

Safety, failure, and incident reports are common artifacts across various domains, including aviation, construction, and wildfire response. These reports are often mandatory to submit, resulting in the culmination of large repositories of text-based documents. Repositories contain a wealth of information from past projects that can be used to provide insight during system development, such as informing risk identification (Gill, Garcia, & Vaughan, 2005). However, it is frequently reported that these repositories are under-utilized. For instance, an audit performed on NASA's Lessons Learned Information System (LLIS) in 2012 found that the repository was not often consulted because the system was "not user-friendly" in addition to other reasons, such as policy and lack of clear objectives for the system (OIG, 2012). Like the LLIS, repositories of text-based documents are often queried by users only via basic search functions and incomplete metadata, making it difficult to quickly extract or retrieve safety-relevant information.

Natural language processing algorithms facilitate a multitude of knowledge management processes, including information extraction and retrieval, at a fraction of the time required for manual processing. However, highly technical language in engineering documentation has been a barrier in applying these tools to engineering domains. While there is a growing body of work on the application of state-of-the-art NLP techniques to engineering documents (NASA, 2022; Sexton & Brundage, 2019), there are few software toolkits focused specifically on extracting and retrieving information useful for studying or improving system safety. Hence, there is a need to develop a system for managing information using advanced methods, specifically for safety engineering applications.

In this research, we present the Manager for Intelligent Knowledge Access (MIKA) toolkit and demonstrate its capabilities using a case study of National Transportation Safety Board (NTSB) reports. The MIKA toolkit currently has two primary capabilities: knowledge discovery (KD) and information retrieval (IR); however, additional capabilities are under development. MIKA uses state-of-the-art algorithms for knowledge discovery and information retrieval, specifically Bi-directional Encoder Representations from Transformers (BERT) and Sentence-BERT (SBERT) models. From leveraging these advanced algorithms, the MIKA toolkit provides customizable methods for engineering-centric knowledge management and access applicable to a user-defined dataset. The MIKA toolkit moves towards an integrated, assistive tool, supporting knowledge access that can easily apply algorithms and analyze data in new datasets in support of a more data-driven systems engineering process, in line with other recent concepts (Kenett, Zonnenshain, & Swarz, 2020). Hence, safety reports from current and prior operations can be analyzed via MIKA to validate early risk identification and assessment for a novel system, as shown in Figure 1. In this paper, we provide relevant background information, an overview of the methods available in MIKA, and apply MIKA in a detailed case study of NTSB reports.

Figure 1: An example of MIKA toolkit's usage in relation to the systems engineering process, adapted from (Shortell, 2015).



Background

To provide context on the MIKA toolkit, in this section we discuss natural language processing, natural language processing in systems engineering, and related tools.

Natural Language Processing

Natural language processing first originated in the late 1900s, where the subfield of semantics aims to understand meaning, implication, and logical conclusions from sentences or text. In the early 2000s, feed forward neural networks were first used for language. During the 2010s, neural networks of all kinds (e.g., convolutional, recurrent, Long-Short-Term Memory, Gated Recurrent Unit, etc.) were heavily used in NLP (Khurana, Koli, Khatter, & Singh, 2022), leading to the development of attention mechanisms, which identify important parts of input data, and later transformers, which learn based on dependency of previous words. In 2018, the Bidirectional Encoder Representations from Transformers (BERT) model was developed (Devlin, Chang, Lee, & Toutanova, 2018), introducing the use of heavily pre-trained models in NLP (Qiu, et al., 2020; Otter, Medina, Medina, & Kalita, 2020). BERT and subsequent derivative models have become the state-of-the-art for numerous NLP tasks, including named entity recognition, summarization, and question-answering (Otter, Medina, Medina, & Kalita, 2020). BERT models are pre-trained on millions of documents for masked language modeling (i.e., masking or removing words to predict the missing word) and next sentence prediction, both of which improve the model's ability to capture context of words. A base pre-trained model can then be fine-tuned for specific tasks, such as document classification and information retrieval. BERT models have been successfully applied to biomedical text (Lee, et al., 2019), financial documents (Araci, 2019), and legal documents (Chalkidis, Fergadiotis, Malakasiotis, Aletras, & Androutsopoulos, 2020). Due to the versatility of pre-trained BERT models and state-of-the-art-performance, BERT is the backbone for multiple MIKA toolkit capabilities and is well suited for NLP tasks in systems engineering due to its ability to be fine-tuned for domain-specific applications (Lee, et al., 2019; Araci, 2019; Chalkidis, Fergadiotis, Malakasiotis, Aletras, & Androutsopoulos, 2020).

Natural Language Processing in Systems Engineering

Research investigating the intersection of NLP and engineering has resulted in a variety of methods in use to assist in different components of the systems engineering process. For instance, the Formal Requirements Elicitation Tool (FRET) leverages natural language processing to assist users in writing complete, consistent requirements (Giannakopoulou, et al., 2020). Another NLP tool used during the requirements phase is the Lessons Learned bot, which retrieves documents relevant to any specified requirement (NASA, 2022). A failure taxonomy extracted using topic modeling can inform design time decisions. Similarly, function knowledge extracted from project documentation can be used to support design time modeling activities (Cheong, Li, Cheung, Nogueira, & Iorio, 2017). During operation, document classification has been used to determine the category of aviation safety reports (Amin, Yother, Johnson, & Rayz, 2022) and question answering has been used to identify information within a specified document (Kierszbaum & Lapasset, 2020). Patterns in aviation accidents have been analyzed using text mining, along with factors that predict fatal accidents, which can be used to improve the safety in future operations (Bazargan, Johnson, & Vijayanarayanan, 2013). Along the same lines, MIKA toolkit capabilities use advanced NLP to intelligently manage documents produced during operations and output information useful for system verification and validation.

Related Tools

There is a large body of products available for knowledge management in business use cases, such as Shelf (Gemshelf Inc., 2022), Nice (Nice, 2022), Watson Explorer (IBM, 2022), and Accenture's intelligent knowledge management solutions (Accenture, 2022). While these products are widely successful and easy to use for business applications, none are open source or designed specifically for engineering documents. It has been argued that highly technical documents often require specialized methods (Han, Sarica, Shi, & Luo, 2021), such as BERT models that have been fine-tuned on engineering documents. There are other tools available that are geared towards engineering documents but focus only on one capability. Nestor, developed at NIST, is one tool that is specifically intended for engineering documents and is used to assist in tagging text documents for computation (Sexton & Brundage, 2019). Perilog, developed at NASA, provides contextual search capabilities (McGreevy, 2005), Although Perilog overcomes shortcomings of conventional search engines (e.g., bag-of-words limitations), the search capability is not sufficient for intelligent knowledge management and the algorithm behind the tool uses pure logic alone, rather than semantic natural language processing methods. Building on the successes of these tools, MIKA provides multiple capabilities packaged in such a way that allows rapid exploration of a data set with multiple tools, reducing the barriers to accessing knowledge stored in the repository. Additionally, MIKA is geared towards system safety, which requires specialized capabilities (i.e., custom named-entity recognition for failure information, fine-tuned models for custom search, etc.). A summary of available tools is provided in Table 1.

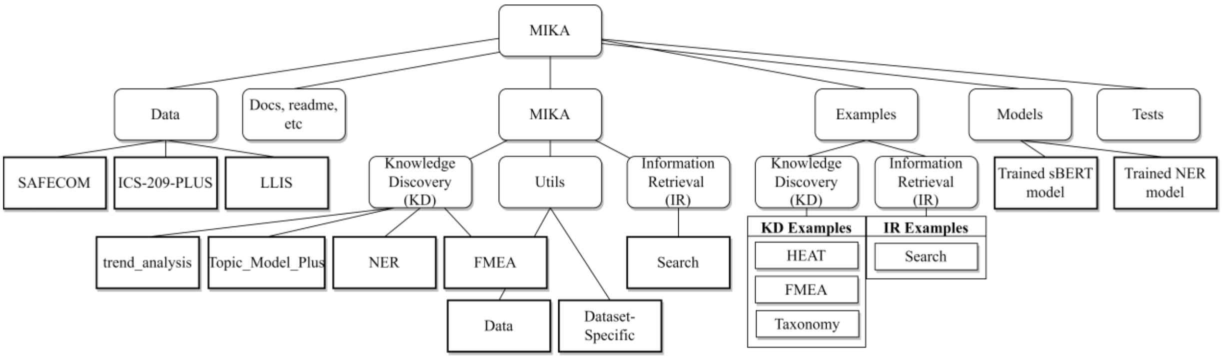
Table 1: Existing knowledge management products and tools.

Product	**Capabilities**	**Availability**	**Use-case**
Shelf	Analytics, search, tagging, content maintenance, question and answering, recommendations	Knowledge management and enterprise packages are available for monthly fees (https://shelf.io/)	Internal organizational knowledge management, external digital assistants
Accenture Intelligent Knowledge Management	Knowledge harvesting, knowledge graphs, taxonomies, semantic search	Only available via consulting contracts, no specific products and pricing (Accenture AI)	Solutions are implemented to customers depending on the business needs
Nice	Question answering, analytics, knowledge management	Pricing dependent on products chosen, limited free trials available (https://www.nice.com/products/digital-self-service)	Custom chat bot for customers to interact with based on the company's knowledge base
Watson Explorer (WEX)	Explore (content mining, search), analyze (analytics), advice (suggestions)	Free community version available, advanced versioning available for cost (https://www.ibm.com/docs/en/watson-explorer)	Intended for mining business insights, but has been used in safety engineering documents (Feldman, Barshi, Smith, & Matthews, 2021)
Perilog	Contextual search	Available for US Government purpose (https://software.nasa.gov/software/ARC-15310-1)	Search applications in knowledge management applied to data such as ASRS
Nestor	Assists annotation of technical text documents	Publicly accessible (https://www.nist.gov/services-resources/software/nestor)	Tagging engineering documents for computation

MIKA Toolkit Methods

The MIKA package is currently organized into three main submodules with future submodules under development. Existing submodules are utilities, knowledge discovery, and information retrieval. In addition to the core package module, the MIKA toolkit also contains a directory of examples, pre-trained models, and tests. Figure 2 provides an overview of the toolkit architecture.

Figure 2: Overview of MIKA toolkit architecture, including the main package module and examples.



Utilities

MIKA includes a utilities submodule mostly comprised of the data class, which is a class for creating data objects that store a data set for ease of working with various MIKA capabilities. Data objects can load a raw or preprocessed dataset, prepare a dataset by dropping empty values or combining columns, preprocess a dataset, and save a dataset to a specified file. Optional preprocessing steps include stop word removal, tokenization, removal of quote marks, spelling correction, lemmatization, n-gram formation, and the removal of short documents with less than a specified number of words. Knowledge discovery and information retrieval capabilities are intended to be applied to a data object to allow users to apply any methods with ease. There are also separate files containing utilities for four datasets (LLIS, SAFECOM, SAFENET, and ICS) used previously to develop MIKA capabilities.

Knowledge Discovery

Within the knowledge discovery submodule, there are four main files: topic model plus, trend analysis, failure modes and effects analysis (FMEA), and named entity recognition (NER). The topic model plus file provides a class definition used to instantiate topic modeling objects. Trend analysis functions can be used to further analyze datasets after topic modeling. When paired together, topic model plus and trend analysis functions can be used to perform hazard extraction and analysis of trends (HEAT). The NER file provides useful functions for training a custom named entity model given annotated data. In the FMEA file, a class is defined for constructing FMEAs using a trained custom named-entity recognition model. Each file and the corresponding capabilities are discussed in the following paragraphs.

Topic Model Plus

The goal of the topic model plus class is to enable a user to easily apply a variety of topic modeling algorithms on the same text data set with multiple sections or columns and save the results in a re-usable and interpretable manner. Three primary types of topic modeling are possible through topic model plus, which are latent Dirichlet allocation (LDA) (Blei, Ng, & Jordan, 2003), hierarchical latent Dirichlet allocation (hLDA) (Blei, Griffiths, Jordan, & Tenenbaum, 2004), and BERTopic (Grootendorst, 2022). Each topic model has different benefits, drawbacks, and use cases. Specifically, LDA requires a user to specify the number of topics in the corpus, whereas BERTopic does not. Similarly, hLDA requires users to specify the number of levels of hierarchy for the model. In LDA and hLDA each document is considered a mixture of topics, whereas by default each document is only assigned one topic with BERTopic; however, MIKA allows a user to input a probability threshold which will allow a document to be considered a mixture of topics. With this in mind, BERTopic is useful for most applications, but in particular is relevant for datasets where not all documents are relevant or failure related. In cases where hierarchal topics are desired, such as when building a taxonomy, hLDA or BERTopic are recommended instead of the flat model provided by LDA.

Topic model plus includes results saving methods that are consistent across topic modeling methods. Results saving options include only topic information (i.e., topic numbers, words, and document counts), as well as separate spreadsheets for a topic taxonomy, topic coherence, topic diversity, and document topic distribution. Users can save either individual views of results, or a collective of all results. For example, a user can save an excel sheet with topic numbers, words, documents, and the best document for the topic for each LDA, hLDA, and BERTopic model. In addition to tabular results, topic model plus also leverages existing visualization methods to produce flat displays for both LDA and BERTopic and hierarchical displays for hLDA and BERTopic. The primary benefit of topic model plus within the MIKA toolkit is the ability to easily run a variety of topic modeling algorithms on the same dataset with minimal code and consistent results for easy comparisons.

Trend Analysis

After extracting hazards or failure modes from text, quantitative trends in hazard occurrence can be analyzed using the trend analysis module. The primary components of MIKA's trend analysis revolve around quantitative methods and include FMEA-style tables, risk matrices, and graphs of both categorical (e.g., hazard region) and continuous (e.g., frequency, severity) data. MIKA also provides functions for statistically analyzing these trends with little effort from users. Together, these functions allow users to visualize trends and determine whether any trends are statistically significant in order to identify important predictors for hazards.

The quantitative FMEA-style table typically includes hazard categories, hazard names, rate, frequency, and severity, where the severity calculation may vary depending on the information available in the dataset. In contrast to a detailed FMEA, which includes failure causes, modes, effects, recommendations, and control processes, trend analysis tables instead focus on high-level hazards and metrics associated with the hazards. Given topic modeling has been performed, a user can identify hazards, corresponding hazard words, and corresponding topics according to the provided hazard interpretation template which allows hazards to be extracted from all documents. This produces various outputs, including the documents associated with each hazard, the relevant hazard words appearing in each document, and the frequency of each hazard. The documents known to describe a hazard can then be used to calculate other metrics depending on the available metadata, such as hazard severity or operational time of occurrence. These outputs can also be used to produce visualizations, including word clouds of the most common words describing each hazard. Graphical trend analysis can be easily performed to produce time series graphs, averages graphs, and categorical data pie charts of relevant hazard metrics and predictors. If a user defines severity values for each hazard, then hazards can be plotted on a risk matrix according to FAA guidelines (FAA, 2017). In addition to graphic analysis, the trend analysis module also provides functions for inferential statistical analysis to determine significant predictors of hazard trends, including correlation matrix, single or multiple linear regression analysis for feature importance, and chi-squared independence tests.

Named-Entity Recognition for Failure Modes and Effects Analysis

While the trend analysis capabilities are mostly focused on quantitative analysis, qualitative analysis can also be performed in MIKA via the FMEA class. Specifically, the FMEA class uses a custom named entity recognition model to detect failure modes, failure causes, failure effects, control processes, and recommendations documented in mishap reports, and thus provides much more detailed qualitative information than the trend analysis methods. In case a user would like to train their own custom named entity recognition model, rather than use the default MIKA model, the NER file provides useful functions. To use the FMEA capability, users must first load a dataset of mishap reports then apply the custom NER model to extract FMEA entities. Then, FMEA rows are formed by either using manual mapping or a clustering algorithm. After the rows of the FMEA are formed, each row is assigned a likelihood according to the rate of occurrence (mishaps per year), severity score according to a user defined severity function, and risk score calculated as the product of severity and likelihood. The FMEA is then automatically post processed for readability, including removing non-words from results and reducing the number of words per cell. As a result, the FMEA class allows a user to input a set of mishap reports and output an FMEA table describing failure modes, failure causes, failure effects, control processes, recommendations, likelihood, severity, and associated risk.

Information Retrieval

MIKA's information retrieval submodule includes the search and custom IR classes and uses data loaded via the Data utility (i.e., users can switch from KD to IR with no additional processing). To perform information retrieval via a search, users must select an SBERT model, which can be downloaded from Huggingface or defined by the user. While MIKA allows a user to fine-tune a pretrained model using its custom IR model class, a default fine-tuned model trained on synthetic query-result pairs (Nogueira & Lin, 2019) is provided. After selecting a model and loading data, sentence embeddings are computed and saved for the document corpus using the selected model. Once sentence embeddings have been either computed or loaded, the built-in search function runs given a query and number of documents to return. The semantic search is performed by computing the cosine similarity between the corpus embeddings and the query embedding. Since semantic search with SBERT (and most modern search methods) return ranked results rather than a finite set of results, the user defines the number of documents to return – it is possible to set this value equal to the number of documents in the corpus to rank all documents, but this is not recommended due to computation time and diminishing usefulness of lower-ranking results. Returned results specify the hit indices, cosine similarity scores, and document text. Returning the scores allows the user to compare the performance of different models or perform other analyses if desired as well as gives the user an indication of how closely the result matches the query. A full description of the information retrieval method used in MIKA has been previously published and is based on the methods of Reimers and Gurevych (2019).

Case Study

To demonstrate the capabilities of the MIKA toolkit, we perform a case study using National Transportation Safety Board (NTSB) aviation accident reports (National Transportation Safety Board, 2022). Each core capability of MIKA is applied to the dataset, including both knowledge discovery and information retrieval. To curate this dataset, we downloaded all datasets from the NTSB website and combined them into one file. The entire dataset includes over 83,000 reports spanning 1948 to 2022. However, we only use reports from 2011-2021 in this case study, resulting in 16,914 documents.

Knowledge Discovery

Failure Taxonomy

A failure taxonomy is constructed in this case study by applying topic model plus to two separate text sections in NTSB documents: accident narrative and cause negative. Hence, the taxonomy pairs co-occuring accident topics with cause topics, resulting in a total of 1,154 rows. In this case study, we use BERTopic to extract topic words from documents. A subset of results with raw topic model outputs is shown in Table 2. The first taxonomy row has failures caused by landing gear, with specific accident outcomes including issues with the nose landing gear, hydraulics, main landing gear, and tire gear. The second row includes failures caused by fuel and oil impacting engine operations, with accident topics detailing mishaps with specific mechanical components and the fuel tank. Finally, the third row includes accidents caused by aborted takeoffs and includes two accident topics about the aircraft weight and a ground collision involving the wings.

Table 2: Subset of the failure taxonomy generated from NTSB reports using BERTopic from Topic Model Plus.

Cause Narrative	Accident Narrative	# of Reports
abort, abort takeoff, decision abort, decision abort takeoff, delayed decision, delayed decision abort, pilots delayed decision, pilots delayed, takeoff, delayed	takeoff, weight, runway, gross, airplane, maximum, gross weight, end, pilot, flaps	7
	taxiway, taxi, airplane, taxiing, wing, pilot, reported, left, truck, right	1
fuel, oil, starvation, fuel starvation, power fuel starvation, engine, loss engine, loss engine power, engine power, power	oil, engine, connecting, rod, connecting rod, crankshaft, bearing, cylinder, revealed, examination	33
	fuel, tank, engine, fuel tank, power, tanks, pilot, gallons, airplane, flight	62
main landing, gear, main landing gear, landing gear, main, right main landing, right main, left main, left main landing, collapse	gear, landing gear, landing, nose, main landing, hydraulic, main landing gear, main, nose landing, nose landing gear	52
	gear, landing, landing gear, mlg, main landing, main landing gear, main, left, fatigue, right	63
	tire, main, left, main landing, main landing gear, landing, landing gear, gear, landing gear tire, gear tire	2

Failure Modes and Effects Analysis

A failure modes and effects analysis-style output is produced from the reports using the previously fine-tuned custom NER model available in MIKA. The resulting FMEA contains seventy-one rows from corresponding mishap types. A subset of the FMEA focusing on six mishaps is available in Table 3. Prior to applying the model, the separate text sections are combined into one narrative for each report. Failure cause, mode, effect, control process, and recommendations in Table 3 are words and phrases directly pulled from reports using the custom NER model and are minimally post processed. FMEA rows are grouped together and operational phases for each row are identified using the mishap type and phase code metadata in NTSB reports. Corresponding likelihood, severity, and risk levels are calculated for each mishap as defined by the FAA (FAA, 2017). Specifically, likelihood categories are assigned according to the rate of mishap occurrence, severity is calculated according to damage and injuries, and risk is calculated according to the combined likelihood and severity categories. (FAA, 2017). Loss of control in flight has the highest risk level, while midair collisions have the lowest risk level due to a low likelihood.

Table 3: FMEA generated from NTSB reports using custom named-entity recognition model. “L” is the likelihood category, “S” is the severity category, and “R” is the risk defined as the product of likelihood and severity.

Phase	Cause	Failure Mode	Effect	Control Process	Recommendation	L	S	R
Bird strike								
Maneuvering; Taxi-from Runway; Takeoff; Landing; Other	at low, bird, surface, a, red - tailed, hawk, female	in - flight, collision, with a, bird, large, pass, foul	bird, substantial, damage, large, broke the front center, windshield	was, towed to the gate, a, bird ingestion and containment	see the public, enter, keep an, eye, on, land	4	3.01	12.05
Ground collision								
Standing; Pushback/Towing; Taxi; Takeoff; Prior to Flight; Taxi; Landing	inadequate visual, lookout, failure, zodiac, did not see the	to see and avoid the, zodiac was, from the, propeller	hit, rear, vertical stabilize, were, destroyed, strike, minor, damage	radio, calls, position, reports, checked the tow, bar, bypass pin	inform, use caution, hold, pilot, clearance, upon, as soon	4	2.88	11.51
Loss of control in flight								
Approach; Descent; Enroute; Landing; Enroute; Takeoff; Taxi; Maneuvering; Initial Climb	pilot's, fatigue, pilot, departed, night, testing, limited, samples, were last	pilot's loss of, entered a, descending left, failure	airplane, spiraled, down, impacting an open, the, lake, damage	intermediate, stops, radar track, toxicology, determination, all, were located at	conform to its type, certificate, documentation, perform, procedures, reset	5	3.66	18.30
Loss of control on ground								
Approach; Descent; Enroute; Prior to Flight; Landing; Maneuvering; Takeoff; Taxi; Initial Climb;	pre - existing crack in the weld, joint, construction of	failure of the left main landing gear, left main	structural, damage, to the, fuselage, right, collapsed, right wing	right aileron, point, applied right rudder to, correct, walked around	remain, reduce, land to the, north, take off to	5	3.04	15.22
Midair collision								
Maneuvering; Taxi-from Runway; Takeoff; Landing; Approach; Emergency Descent	failure of both pilots to see and avoid the other	midair, collision, airplanes, collided, collided nearly head -, on, main	impacted the, terrain, a utility, pole, to, rest, an, immediate	wingtip, mounted strobe anticollision, lights, graphic, remote communications,	alter, course, pass well, not have passed over, under	3	3.40	10.19
Turbulence encounter								
Takeoff; Landing; Descent; Approach; Maneuvering	with, of, cumulonimbus, clouds, slowly, activity, temperature, strong, turbulence was	and, turbulence, tailwind, knots, tailwind suddenly, decreased, airplane began to	indicated speed of the airplane quickly, increased, the, airplane, passenger	captain activated the speed, brakes, disengaged the autopilot, crew, turned	expect light - to - moderate, before, cabin	4	3.79	15.17

Trend Analysis

Trend Analysis is performed for each mishap of interest from the FMEA in Table 3 and provides insight into trends in mishap occurrence. For the trend analysis, time series of hazard frequency and severity are graphed, as well as average mishap severity, the distribution of mishaps over activated/non-activated flight plans, and the distribution of mishaps over sky conditions. In this case study, mishap frequency and severity trends are tracked over the ten-year time period in Figure 3(a) and 3(b). The occurrence frequency of midair collisions, ground collisions, loss of control in flight, and turbulence encounters appears to decrease from 2019-2021; however, loss of control on the ground and bird strikes do not seem to decrease over time. The severity of bird strikes, midair collision, and turbulence have more variance over the years, while ground collisions and loss of control both inflight and on the ground have more consistent severity. From Figure 3(c), on average, midair collisions and loss of control in flight are the most severe mishaps, followed by turbulence, loss of control on the ground, bird strikes, and ground collisions. Figure 4 shows the distribution of mishaps across categorical metadata, with Figure 4(a) showing mishaps over flight plan activations and 4(b) showing mishaps over sky conditions. Midair collisions and loss of control in flight have a greater proportion of reports with no activated flight plan when compared to turbulence encounters and bird strikes, while the majority of turbulence reports were from flights using a flight plan. Birdstrikes, loss of control on the ground, and ground collisions have a larger proportion of occurrences in sky conditions with clouds when compared to the other mishaps, which is confirmed using chi-squared tests ($p<0.001$). The mishaps are placed on a risk matrix in accordance with FAA Order 8040.4B from 2017 in Figure 5. From the risk matrix, it is clear that all six mishaps are in the unacceptable high risk category as defined by the FAA. While individual flights may be deemed low risk for a certain mishap, the analysis provided by MIKA demonstrates that historically mishaps in aviation are still at too high of a risk level.

Figure 3: Times series graphs of mishap trends in severity (a) and frequency (b), with average severity and standard deviations in (c).

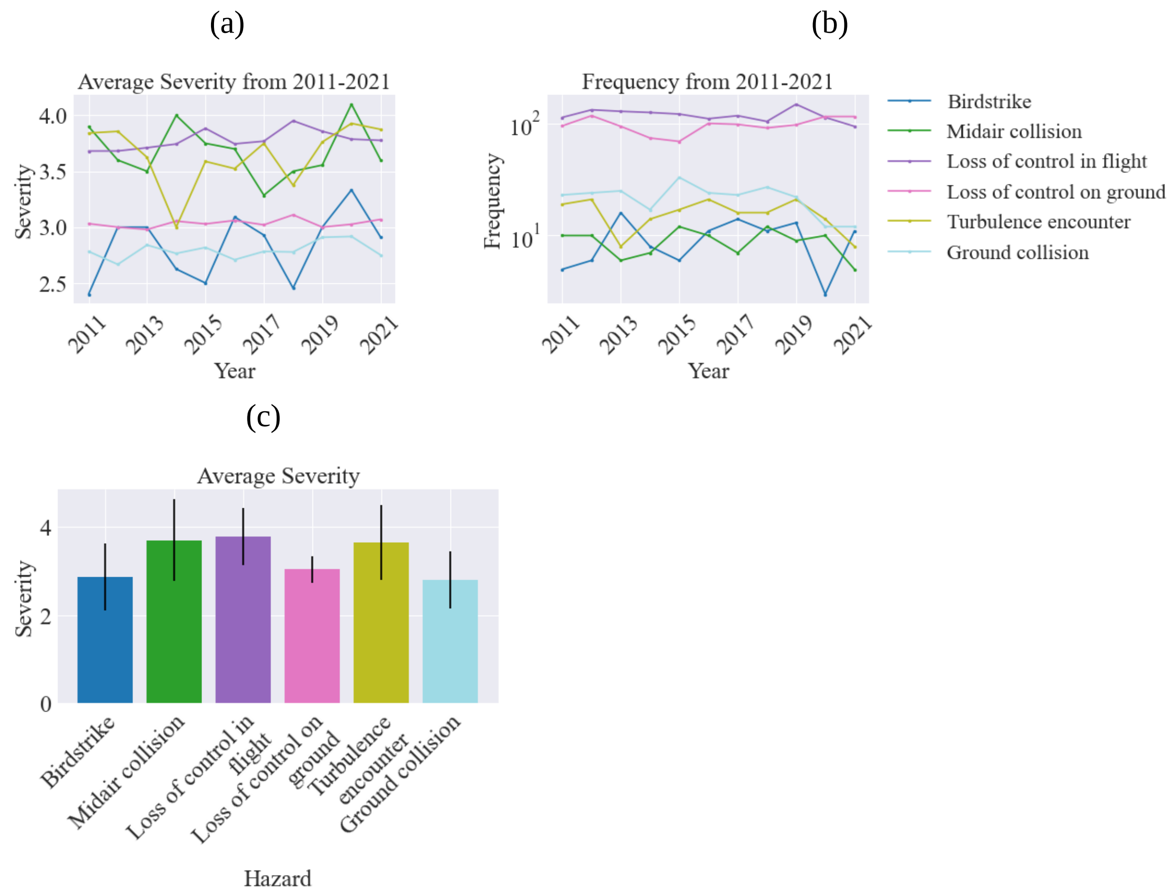


Figure 4: Distribution of mishaps over activated flight plans (a) and sky conditions (b).

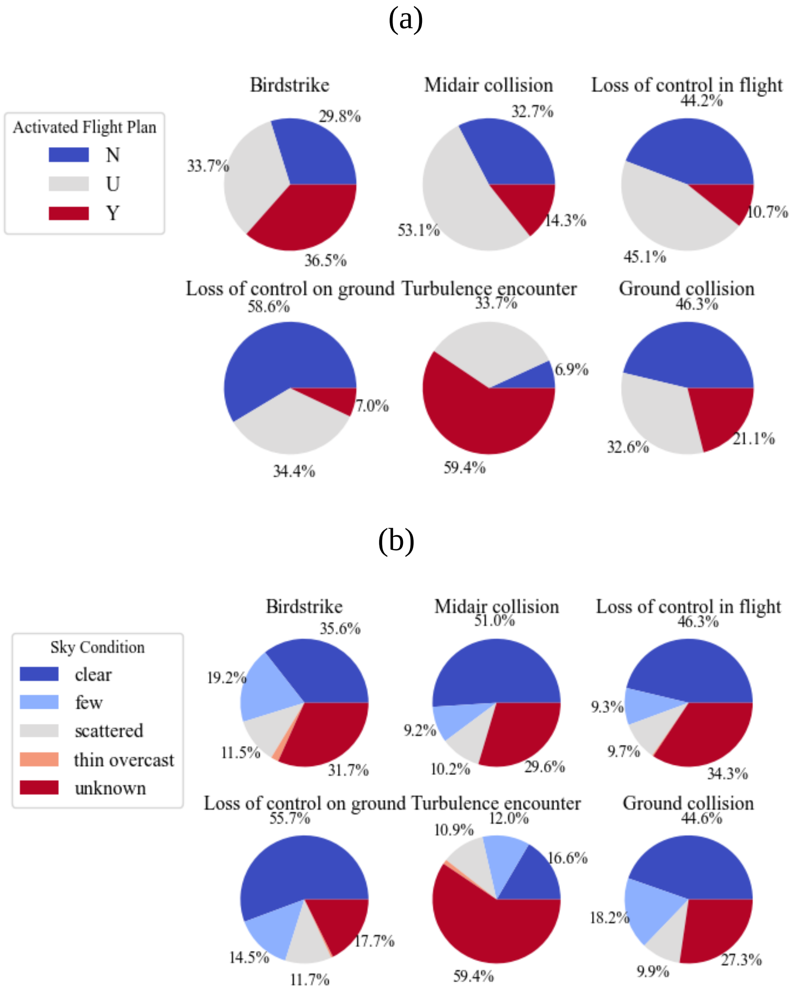
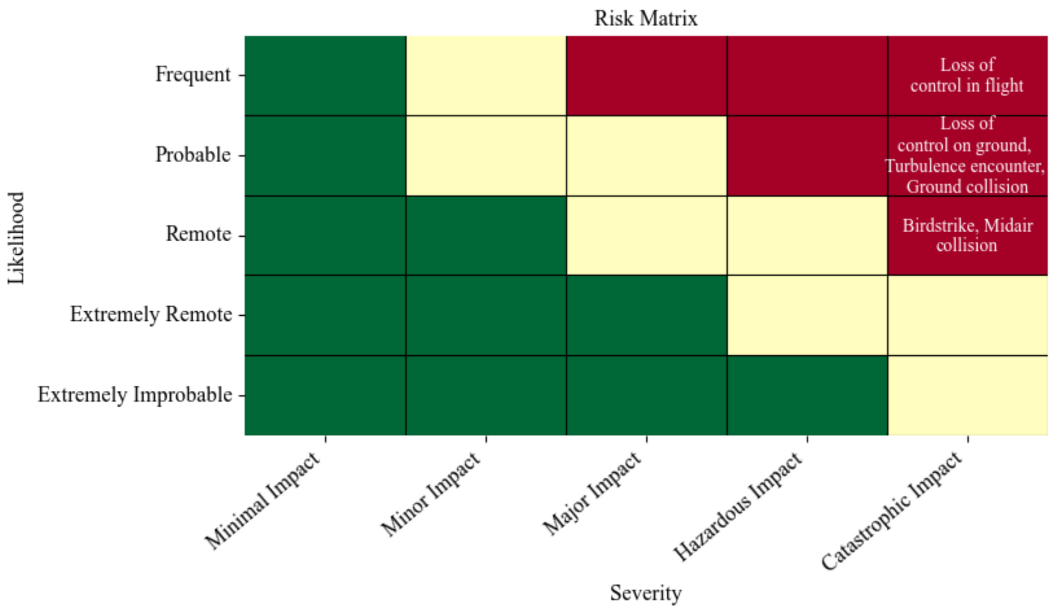


Figure 5: Risk matrix with selected mishaps placed according to FAA Order 8040.4B.



Information Retrieval

For the case study, an sBERT model previously fine-tuned using documents from NASA's Lessons Learned Information System (LLIS) is used to search the NTSB documents. Three queries are tested: "what components are vulnerable to fatigue crack", "what are the consequences of a fuel leak", and "what are the risks of low visibility". These queries are selected based on familiarity with the reports available in the NTSB, likely applicability to future operations (i.e., they are not overly specific), and structure of the training data (short query). Results are shown in Table 4. The top three documents returned are shown along with the result's similarity score and a manually selected relevant text excerpt. Based on similarity, the third query results in the least relevant results, though qualitatively the results are at least adjacent in subject matter. The method returns ranked results, not a finite set, so it is possible that for queries with few relevant results, some irrelevant results may be returned.

Table 4: Search results from querying NTSB documents using an sBERT model fine-tuned on NASA's LLIS.

Result No.	Document ID	Similarity	Relevant Text Excerpt
Query 1: what components are vulnerable to fatigue crack			
1	20121204X63622	0.746134	…one of the first-stage compressor blades had fractured due to fatigue cracking…
2	20141027X62323	0.728395	…the crankshaft had fractured due to a fatigue crack that had initiated at a fillet radius…
3	20160114X73526	0.715234	… forward wing spar likely fractured due to compression loading from wing loads…
Query 2: what are the consequences of a fuel leak			
1	20120129X84717	0.756848	Witnesses and fire department personnel noted fuel leaking due to a cracked fuel line…
2	20140812X10218	0.723396	…he removed the fuel tank caps and observed no fuel in the left fuel tank…
3	20110128X23759	0.705061	…examined the airplane and determined the left and right main fuel tanks were empty…
Query 3: what are the risks of low visibility			
1	20110105X11652	0.562436	…during periods of low visibility, the supporting senses sometimes conflict with what is seen …
2	20110131X11636	0.526502	…during periods of low visibility and night conditions, the supporting senses sometimes conflict with what is seen…
3	20160921X20513	0.518302	…pilot of the float-equipped airplane encountered low visibility due to ground fog.

Discussion

In this case study, MIKA's capabilities enable rapid discovery and exploration of the NTSB dataset. Prior to applying MIKA, the authors had limited familiarity with the dataset and were able to discover multiple themes and retrieve reports relevant to given information needs, thereby increasing familiarity more quickly than would be possible with more limited tools (e.g., searching via metadata or keywords). In turn, MIKA uncovered information on aviation accidents present in current day operations, including identifying the high level of risk associated with ground collisions, midair collisions, bird strikes, loss of control, and turbulence. Foundational knowledge about aviation mishaps can be discovered through MIKA, which can then inform the future concept of operations and design of mitigations and advanced safety-oriented systems. Information discovered using MIKA, including hazard rates, severities, and effects, can be used to validate and inform requirements. While more rigorous studies on the use of MIKA within the systems engineering process will be necessary, this application provides an initial indication of MIKA's value to systems engineers in addition to a demonstration of its capabilities. Application of the toolkit to the NTSB dataset demonstrates the generalizability of the toolkit and ease of implementation, which is important to MIKA's stated goal of reducing barriers to the application of its NLP-based capabilities in engineering projects. Leveraging historical datasets is expected to be particularly useful in large-scale, complex, safety-critical systems in which it is important, and difficult, to anticipate possible failures.

Conclusions and Future Work

This paper presents the MIKA toolkit, which provides capabilities for knowledge discovery and information retrieval that utilize state-of-the-art natural language processing techniques packaged into a single, integrated tool. In this paper, MIKA's capabilities are described from a user-oriented perspective, the algorithms used to support these capabilities are detailed, and the usage is demonstrated on a set of reports from the NTSB. MIKA enables a user to quickly apply multiple techniques, including topic modeling, named-entity recognition, trend analysis, and sematic search, in order to analyze a set of documents and extract information relevant to systems engineering. MIKA is intended to improve an organization's ability to leverage knowledge stored in documentation from past projects and accidents, an often underutilized source due to the challenges involved in managing and finding relevant information in a large dataset. Effectively, this means reducing the space of unanticipated hazards and problems by more reliably anticipating those problems that have appeared in past projects and operations. MIKA is designed to be integrated into the systems engineering process flexibly at multiple junctions such that unanticipated problems can be most effectively reduced through enhanced knowledge management.

Future work will further develop the MIKA toolkit to support new and improved capabilities (e.g., error checking) for intelligent knowledge management as they are developed. Additionally, a user interface will be developed to improve usability and accessibility of the tool. Finally, workflows will be developed for integrating MIKA with other related verification and validation tools for safety assurance. There is significant potential in using MIKA to support early design modeling activities, the formulation of a safety case, and other activities in systems engineering for which excellent tools already exist. For example, a MIKA-enabled modeling dashboard could assist validating the completeness of modeled scenarios. Providing a convenient interface between these tools and MIKA would support the adoption of MIKA as well as other tools into systems engineering activities.

References

Accenture. (2022, November). Retrieved from Intelligent Knowledge Management Solutions: https://www.accenture.com/us-en/services/applied-intelligence/intelligent-knowledge-management

Amin, N., Yother, T. L., Johnson, M. E., & Rayz, J. (2022). Exploration of Natural Language Processing (NLP) Applications in Aviation. *Collegiate Aviation Review International, 40*(1), 203-216.

Andrade, S. R., & Walsh, H. S. (2021). Wildfire Emergency Response Hazard Extraction and Analysis of Trends (HEAT) through Natural Language Processing and Time Series. *IEEE/AIAA 40th Digital Avionics Systems Conference (DASC).*

Andrade, S. R., & Walsh, H. S. (2022). What Went Wrong: A Survey of Wildfire UAS Mishaps through Named Entity Recognition. *2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC).*

Andrade, S., & Walsh, H. (2022). Discovering a Failure Taxonomy for Early Design of Complex Engineered Systems Using Natural Language Processing. *Journal of Computing and Information Science in Engineering (JCISE).*

Araci, D. T. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained. *arXiv*.
Bazargan, M., Johnson, M., & Vijayanarayanan, A. (2013). *An Evaluation of AIRES and STATISTICA Text Mining Tools as Applied to General Aviation Accidents.* Washington, DC: Department of Transportation.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning Research*, 993-1022.
Blei, D., Griffiths, T., Jordan, M., & Tenenbaum, J. (2004). Hierarchical Topic Models and the Nested Chinese Restaurant Process. *Advances in Neural Information Processing Systems*.
Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The Muppets straight out of Law School. *Findings of the Association for Computational Linguistics: EMNLP 2020.*
Cheong, H., Li, W., Cheung, A., Nogueira, A., & Iorio, F. (2017). Automated Extraction of Function Knowledge From Text. *Journal of Mechanical Design, 139*(11), 111407.
Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*. doi:10.48550/ARXIV.1810.04805
FAA. (2017). *Safety Risk Management Policy Order 8040.4B.* Washington, DC: U.S Department of Transportation.
Feldman, J., Barshi, I., Smith, B., & Matthews, B. (2021). Reports of Resilient Performance: Investigating Operators' Descriptions of Safety-Producing Behaviors in the Aviation Safety Reporting System. *42nd International Symposium on Aviation Psychology*, (pp. 122-127).
Gemshelf Inc. (2022, November). Retrieved from Shelf: https://shelf.io/
Giannakopoulou, D., Mavridou, A., Rhein, J., Pressburger, T., Schumann, J., & Shi, N. (2020). Formal requirements elicitation with FRET. *International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ-2020).*
Gill, P., Garcia, D., & Vaughan, W. (2005). Engineering Lessons Learned and Systems Engineering Applications. *43rd AIAA Aerospace Sciences Meeting.* AIAA.
Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
Han, J., Sarica, S., Shi, F., & Luo, J. (2021). Semantic Networks for Engineering Design: A Survey. *Proceedings of the Design Society: ICED21* (pp. 2621-2630). Gothenburg, Sweden: Cambridge University Press.
IBM. (2022, November). Retrieved from Watson Explorer Documentation: https://www.nice.com/products/digital-self-service
Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*. doi:https://doi.org/10.1007/s11042-022-13428-4
Kierszbaum, S., & Lapasset, L. (2020). Applying Distilled BERT for Question Answering on ASRS Reports. *2020 New Trends in Civil Aviation (NTCA).*
Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2019). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. doi:10.1093/bioinformatics/btz682
Lee, M. (2022, July). Tomotopy. Zenodo. doi:10.5281/zenodo.6868418
McGreevy, M. (2005). *NASA", Using Perilog to Explore "Decision Making at.* NASA.
NASA. (2022). Lessons Learned Bot v1.2. *LAR-19934-1*.
National Transportation Safety Board. (2022). Aviation Accident Data. Retrieved from https://data.ntsb.gov/avdata

Nice. (2022, November). Retrieved from Digital Self Service: https://www.nice.com/products/digital-self-service
Nogueira, R., & Lin, J. (2019). From doc2query to docTTTTTquery. *Online preprint*.
OIG. (2012). *Review of NASA's Lessons Learned Information System.* NASA.
Otter, D. W., Medina, J., Medina, J., & Kalita, J. K. (2020). A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems*, 1-21. doi:10.1109/TNNLS.2020.2979670
Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences, 63*(10), 1872-1897. doi:10.1007/s11431-020-1647-3
Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP 2019.*
Sexton, T., & Brundage, M. (2019). Nestor: A Tool for Natural Language Annotation of Short Texts. *Journal of Research of the National Institute of Standards and Technology*.
Shortell, T. M. (2015). INCOSE systems engineering handbook: a guide for system life cycle processes and activities. John Wiley& Sons.
Walsh, H. S., & Andrade, S. R. (2022). Semantic Search with Sentence-BERT for Design Information Retrieval. *ASME 2022 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE).* St. Louis, MO, USA: ASME.

Biography

Sequoia Andrade. Sequoia is a Research Engineer at HX5 on a contract with NASA Ames Research Center, working in the Robust Software Engineering technical area of the Intelligent Systems Division. She earned her B.S. with a double major in Mathematics and Psychology from Santa Clara University in 2020 and is currently pursuing a M.S. in Statistical Data Science at San Francisco State University. In 2022, she was awarded a fellowship from the NSF Graduate Research Fellowship Program. Her research interests include utilizing statistical and machine learning methods for social good applications, such as safety analysis.

Hannah Walsh. Dr. Hannah Walsh is a Computer Engineer in the Intelligent Systems Division of NASA Ames Research Center. She earned her Ph.D. and M.S. in Mechanical Engineering with emphasis in Design from Oregon State University in 2020 and 2018, respectively, and her B.S. with a double major in Aerospace Science & Engineering and Mechanical Engineering from the University of California, Davis in 2016. Her research interests include the application of artificial intelligence to the design process with a particular emphasis on improving safety in complex systems.