Assessing the Use of UAS-Related Terms in ASRS using Seeds for Topic Modeling

Carlos Paradis*
KBR, Inc, California, USA

Misty D. Davies† and Becky L. Hooey‡
NASA Ames Research Center, Moffett Field, CA, 94035, USA

Context: The NASA Aviation Safety Reporting System (ASRS) is a voluntary confidential system that disseminates reports received from personnel involved in aviation operations after de-identifying them. These reports are used by the community to improve overall aviation system safety.

Aim: We propose and execute an experiment to assess the use of seed term topic modeling over the database narratives to identify Unmanned Aircraft System (UAS) reports. The use of seed term topic modeling enables users to identify groups of conceptually similar narratives associated to a topic of their interest.

Method: We use a collection of narratives, expert-selected words, and report metadata that separates UAS from non-UAS reports to assess if seed topic modeling can be used to improve ASRS searches.

Results: For simpler queries, seed topic search observes a higher recall and lower precision than the existing DBOL (DataBase OnLine) search in operation. However, the best results are obtained when seed topic search is used as a search suggestion system to be executed on the DBOL.

Conclusion: Utilizing a combination of both the existing method and the proposed method, users can expand their search vocabulary about subjects of interest while improving the quality of results.

I. Introduction

The NASA Aviation Safety Reporting System (ASRS) is a voluntary confidential reporting system. The ASRS receives reports from pilots, air traffic controllers, flight attendants and others involved in aviation operations. Every ASRS report is screened within three to five days of receipt by an ASRS expert analyst who has over 10 years of experience in the aviation domain. Of interest to us in this step is that reports are also identified as related to Unmanned Aircraft Systems (UAS) if the report was submitted by an UAS operator or if the content of the narrative described UAS operations. We will later describe how we leverage this information for the proposed method evaluation.

The de-identified reports are then disseminated to the aviation community in a number of ways including entry into an online database. The retrieval of these reports is done through word search and metadata search. ASRS uses DBOL (DataBase OnLine), which is standard software that retrieves and updates data on any Integrated Database Management System database. Further augmenting the report search capability in ASRS online database would therefore be beneficial to the community it serves.

Because ASRS report search capability plays a crucial role in the discovery of various aviation topics, this work's goal is to augment the search capability using topic modeling. We considered that, in addition to free text search and metadata search, the use of topic modeling could enable a third means to explore ASRS reports. That is, the use of topic modeling would enable users to explore groups of related reports if their narratives are similar as identified by the topic model. Intuitively, our proposed optional means to explore reports [1, 2] could be seen as an automated way to provide users with additional report sets to explore on the ASRS Report Set page *.

*Technical Professional - Data Science.
†Research Computer Engineer, Intelligent Systems Division, Mail Stop 269-1, AIAA Associate Fellow.
‡Director, NASA Aviation Safety Report System, P.O. Box 189.
*`https://asrs.arc.nasa.gov/search/reportsets.html`

Each identified group of similar reports is also assigned a set of words which are assumed [2] by the model to summarize what the group of reports is about. These words have a similar purpose as the short summaries that the ASRS Report Set page associates to each report set. For example, a group of reports that has topic-model assigned words including 'cabin fire fumes' could be interpreted as equivalent to the manually curated *Cabin, Smoke, Fire, Fumes, or Odor Incidents* report set †.

While the devised framework in previous work provides an automated way to expand report set navigation, a limitation of the existing framework lies in the use of classic topic modeling, which does not allow for users to influence the discovery of topics for subjects of their interest.

In this work, we propose to address this limitation by evaluating the use of seed topic modeling in ASRS. By allowing for seed terms in topic modeling, we can then leverage the words used in a user query to retrieve records of interest using an additional and parallel search method to keyword matching. In this manner, a seed topic modeling presentation of results is now equivalent to that of performing an ASRS query using free text. However, instead of using word matching, the similarity of narratives is used to retrieve related reports instead.

To evaluate our method in this work, we chose the five most-used ASRS user queries that were associated with UAS in general. Because we are able to identify UAS reports after they are screened, we are able to devise an experiment where we can assess the precision and recall of each of the five queries using both the existing query in place, and also the proposed seed topic modeling.

More generally, we define the following research questions:

RQ1 What is the performance of an UAS query using a classic topic model?

If we were to re-purpose our prior work topic model for searching UAS-related topics, how many UAS reports would the user expect to recover?

RQ2 What is the performance of each UAS query by the proposed model system?

Having established a baseline in RQ1 for our prior work, we now investigate the performance of the free text query in ASRS versus the use of a seed topic model for popular UAS queries.

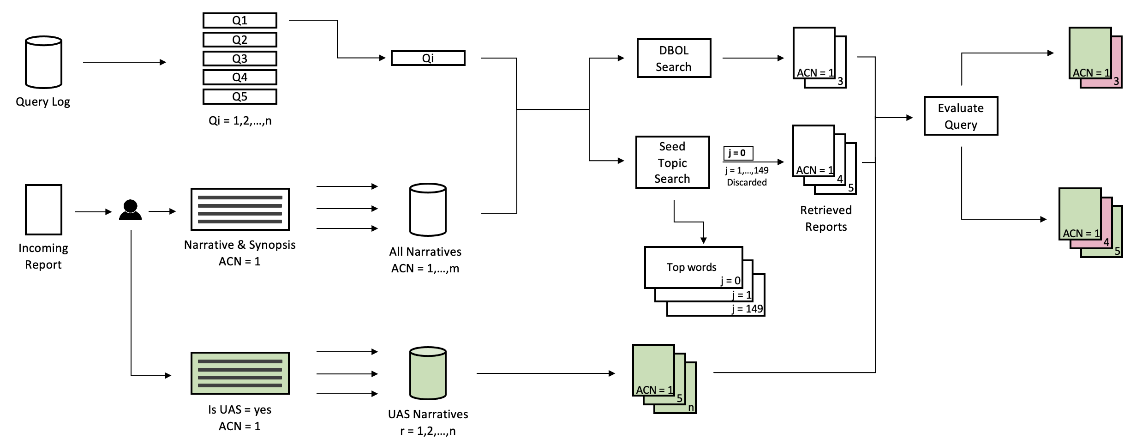
RQ3 Can the suggested list of terms of the topic model be leveraged as term suggestions?

As noted previously, topic models not only group documents, but also associate a list of words summarizing each group. In the context of using topic modeling for a query, an interesting property of the set of automatically-generated words is that they may include words related to UAS but not thought of by the user, and hence could be used in successive searches.

II. Related Work

Related literature in text mining aviation has focused in classification and grouping of reports both in the NTSB [3, 4] and the ASRS [5–9]. For ASRS, the classification in general seeks to establish linkage between the narratives and the metadata fields available on the reports in an automated manner [5, 6]. The closest work to ours is discussed in [6]. In [6], two systems are proposed using topic modeling: The first uses classical topic modeling, and is similar to our prior work [1, 2, 10], which is evaluated in the experimental design of Figure 3. The second method proposed by the authors is similar to our current proposed work, which is presented in the experimental design of Figure 1. In the second method, the authors propose the *timePlot* system, in which a user can input an entire report to search for similar reports. In our experimental design of Figure 1, we use the user query, instead of a input report, to search for related reports. Despite the similarity in the use of topic modeling, our work diverges from [6] in that we empirically evaluate the discoverability of reports in the context of a query, whereas the authors qualitatively evaluate the interpretability of the retrieved topics.

Fig. 1 Seed Topic Search Experiment Setup. What is the UAS reports discoverability?



III. Method

In this section, we first explain how we collected and prepared the dataset, then the algorithms used, and then finally present how we devised the evaluation setup using them. We devised two experimental setups, as shown in Figures 1 and 3. The following dataset subsection is common to both, hence we will use Figure 1 as reference.

A. Dataset

In Figure 1, we can observe on the left that we used three different datasets: 1) Past UAS user queries, 2) the narrative and synopsis field of each report, and 3) domain-expert-generated labels identifying incoming reports as either UAS or non-UAS reports. We will now expand on each of these data sources.

1. Past UAS user queries.

We examined ASRS queries from April 2021 through July 2021 to select representative UAS queries for our experiment. Because the screening of queries had to be done separately and is labor intensive, we chose a smaller time window to choose queries from when compared to the other datasets. Tables 1 and 2 ranks all identified UAS-related queries by frequency from top to bottom in the time period. In Table 1, the queries in bold which are the most frequently general queries performed on ASRS, were used in this work for the use cases. Table2 separates what we deem user intent on obtaining a sub-topic of UAS instead of "any" UAS related report, and is included only for completeness, but not used in the experiments. This distinction on query specificity is important because our domain-expert label dataset shown in Figure 1 only labels UAS reports from non-UAS reports, and therefore we would be unable to use it to assess more specific queries.

2. Narrative and Synopsis of all Reports

As noted in the introduction, every incoming ASRS report is screened within three to five days of receipt by an ASRS expert analyst. This is depicted on the left of Figure 1. In the scope of this work, we consider two data products to result from this step. The first is the narrative & synopsis, and the second, the *domain-expert UAS labels*. We chose a corpus of narratives & synopses from December 2020 to August 2021 due to the availability of UAS labels during that time period.

For both narratives and synopses, we performed some text pre-processing, which helps improve the performance of our subsequent topic search step. Specifically, in addition to common stopwords [‡], the following words were removed from the narratives as they would not convey meaning if occurring among the 'top-n terms': 'x', 'y', 'z', 'xx', 'yy', 'X', 'Y', 'Z', 'ZZZ', 'ZZZZ', 'zzz', 'zzzz', 'zzzzz', 'zzz1', 'zzz2', 'zzz3', 'zzz4', 'zz2', 'zz3'. These are codes used by ASRS analysts to de-identify information such as airports and navigation waypoints.

[†]https://asrs.arc.nasa.gov/docs/rpsts/cabin_fumes.pdf
[‡]https://algs4.cs.princeton.edu/35applications/stopwords.txt

Table 1 Top user performed UAS general queries

General Queries
(DRONE)
(UAV)
(UAS)
(DRONE OR UAS OR UAV)
(UAV OR UAS OR UNMANNED OR DRONE)
(DRONE OR UAS)
(UAS OR UNMANNED AERIAL SYSTEM)
(UAV OR UAS OR DRONE)
(UAV AND UAS)
(DRONE OR UAS OR UAV OR SUAS)
(DRONE OR UAS OR UNMANNED AERONAUTICAL SYSTEM)
(UNMANNED)
(UNPILOTED AERIAL VEHICLE)

Table 2 Top user performed UAS specific queries

Specific Queries
(MATRICE)
(RECREATIONAL)
(LOSS OF GPS)
(BALLOON AND DRONE)
(GEAR UP LANDING)
(DRONE AND FAST)
(DRONE, QUAD)
(QUADCOPTER)
((MIDAIR OR NEAR MISS) UAS OR UNMANNED)
(DRONE% AND (MIDAIR OR MISS))
(STREAMING VIDEO USING MY PERSONAL HOBBY DRONE.)

3. Domain Expert UAS Labels

We consider the obtained Domain Expert UAS labels from December 2020 to August 2021 to serve as our "ground truth". Specifically, and as shown in our experiment setup diagrams in Figures 1 and 3, this ground truth is used to evaluate and compare the discoverability of UAS reports in each of the experimented UAS queries. Both DBOL and Topic searches have no knowledge of whether a report is labeled as a UAS report or not.

In the following subsections we describe the different searches we devised on the two experiments. The DBOL Search and the Seed Topic Search are referenced in Figure 1 and are associated to RQ2 and RQ3, while Non Seed Search defined in Figure 3 is used to answer RQ1.

B. Model Evaluation

To explain in Figures 1 and 3 how we define the Topic Searches, we must first define the evaluation metrics used in our work. These metrics are also used at the final "Evaluate Query" step of both diagrams. For model evaluation, we used three different approaches: Confusion matrices, Precision & Recall, and the F-1 Measure. In the following definition of metrics, TP, TN, FP, FN, are the abbreviation of True Positive, True Negative, False Positive, and False Negative.

1. Confusion Matrix.

For the model comparison and presentation of results (i.e. the "Evaluate Query" step in Figures 1 and 3), we used the confusion matrix. The confusion matrix is more commonly reported in the format shown in Table 3. We prefer the use of confusion matrices here as it allows us to identify the benefits and shortcomings of both search methods. In the presentation of results, we also included the more commonly reported metrics in the literature, precision and recall, as shown in equation 1, and F_1 measure in equation 2.

The F_1 measure is of particular interest for us for model tuning, which is used to define the "Topic Search" step in Figures 1 and 3). Being a single point metric, it facilitates visualizing a large number of model comparisons, which we require when choosing empirically the number of topics as we will discuss in the model tuning section.

Table 3 Confusion Matrix Format.

		Reference		
		No	Yes	Total
Prediction	No	TN	FN	$TN + FN$
	Yes	FP	TP	$FP + TP$
	Total	$TN + FP$	$FN + TP$	N

2. Precision and Recall.

$$Precision = \frac{TP}{TP + FP}; Recall = \frac{TP}{TP + FN} \tag{1}$$

3. F_1-Measure.

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{2}$$

Having defined the common parts of both experimental setups, i.e. the dataset and the evaluation metrics, we now describe the different types of searches used in our experimental setup, which we will present in the results section.

C. DBOL Search

We first present the DBOL (the standard search available online[§]) in Figure 1. The search requires both a given Query of interest Q_i, and a database of narratives and synopses.

To evaluate the existing search method, each of the five queries were performed with the following parameters:

- **Date of Incident:** From December 2020 to August 2021 inclusive.
- **Text: Narrative/ Synopsis:** One of the five queries in bold from Table 1 at a time.
  - **Fields to search:** Narrative and Synopsis boxes were checked.

It is important to note that the ASRS search engine can perform more advanced types of queries[¶], however we limited the search strategy to those more frequently used for UAS searches as previously discussed. Intuitively, we assume an ideal query concerning UAS-related terms would retrieve all UAS-labeled reports from our ground truth dataset (recall), without mistakenly returning non-UAS labeled reports (precision). Achieving the ideal result in this case, however, *is limited by the user's knowledge of the vocabulary contained in narratives that are associated to UAS*, and even if extensive at the time, the vocabulary can also evolve over time. This limitation is not true to topic search, however, as it does not rely solely on exact word match.

[§] `https://asrs.arc.nasa.gov/search/database.html`
[¶] `https://asrs.arc.nasa.gov/search/dbol/strategies.html`

D. Topic Search

Because our proposed Seed and Non-Seed Topic Search shown in Figures 1 and 3 respectively adapt topic modeling to serve as a search system, we will first define topic modeling, and in the subsequent sections explain how their results are leveraged in the context of a search.

1. Topic Modeling

Classic topic modeling algorithms, such as WarpLDA [11] are primarily a grouping algorithm, i.e., at least two groupings (also known as topics) should conceptually be expected to exist in the corpus. Here, our corpus is the ASRS narratives. "Seed topic models" such as Seed-LDA [12] and CorEX [13], which are also grouping algorithms, allow for user-specified words to "lead" the formation of the groupings. While the underlying assumptions of classic and seed topic algorithms varies, one major feature difference in the scope of this work is being able to utilize user specified words (e.g. from a user query).

When words are not specified to a seed topic model, the algorithm functions as a classic topic model. Because a seed topic model can function without seed words, it suffices using only a seed topic model to answer all our RQs. All topic model algorithms discussed in the scope of this work will a) group narratives based on their similarity, and b) associate a set of words to each grouping intended to summarize their meaning. Finally, we must also specify *a priori* the number of topics we expect to find across all narratives.

We chose to use CorEX [13], as this model allowed greater flexibility in choosing which groupings should be influenced by the user words. For example, if the apriori number of topics were three, Seed-LDA [12] would generate one topic *unrelated* to the user words, and two topics related to the user words. In contrast, CorEX [13] allowed us to choose one or more topics to be *related* to the user words. In this manner, as the number of topics increased for the Seed LDA model, the user words would further spread across different topics. Meanwhile, in CorEX, we could control for one topic to be associated to the user words and all other topics to be unrelated to them, regardless of how many topics were chosen. We found the CorEX's feature to isolate in one topic the reports associated to user words ideal for using in a query use case, as we would like to retrieve only the user-specified topic, its associated reports, and a summary of the topic.

2. Deterministic Mapping

Topic modeling provides a probabilistic mapping from narratives to each grouping. Similarly, CorEx provides a soft assignment between narratives and groupings. Because the narrative being UAS-related or not according to its metadata is deterministic, we must establish a deterministic mapping. In our prior work, and to answer RQ1 and RQ2 in this work, we defined a deterministic mapping from narratives to topics using the highest probability of the topic given the narrative in Eq. 3.

$$argmax_{z_i} p(z_i|d) \tag{3}$$

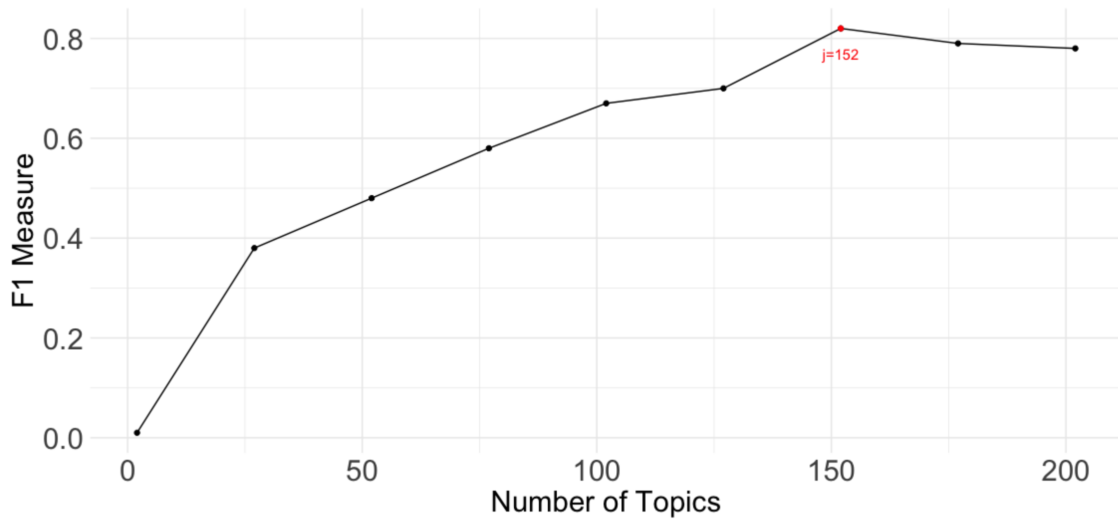
Where z_i is a given topic, and d is a narrative.

3. Model Tuning

To choose the apriori number of topics, we randomly selected one of the five queries to perform model tuning, resulting in the 5_{th} query, Q_5 = (UAV OR UAS OR UNMANNED OR DRONE). We then performed multiple runs of CorEX, varying on each only the number of topics from $j = 0$ to $j = 200$. For each model, the performance was recorded using the F-1 Measure, which provides a summary metric for both precision and recall. Figure 2 provides the model performance for different number of topics j.

We can see the model has best performance when $j = 150$. We thus used $j = 150$ number of topics for all models in this work. While the absence of a separate test dataset to evaluate the choice on the number of topics may lead to more optimistic results of both topic searches, we argue a more sophisticated operational setup could leverage past user queries in combination with available metadata to pre-tune various models depending on the words used in the query. We defer to future work evaluating both a more complex experimental setup and whether the results derived in this work on a test dataset are optimistic.

Fig. 2 Model Tuning using F-1 Measure.



4. Seed Topic Search

Now that we have presented our motivations in using the CorEX model [13], the criteria to choose the number of topics, and how probabilistic assignments were made deterministic to all our three research questions, what remains is defining how the topic models were repurposed as an optional query alternative to the DBOL Search.

First, let's consider seed topic models for seed topic search, as shown in the experimental setup of Figure 1. Reusing seed topic modeling as a search algorithm requires little modification, because CorEX guarantees topic j = 0 is always related to the used seed terms. By using Q_i words as the seed terms, and returning to the user the reports assigned to topic j = 0, while discarding the remaining topics j = 1 through j = 149 inclusive, we can effectively use the seed topic search as a query system. Additionally, the set of words assigned to topic j = 0 can be seen as "recommended terms" for future user queries. This setup is used for both RQ2 and RQ3.

5. Non-Seed Topic Search

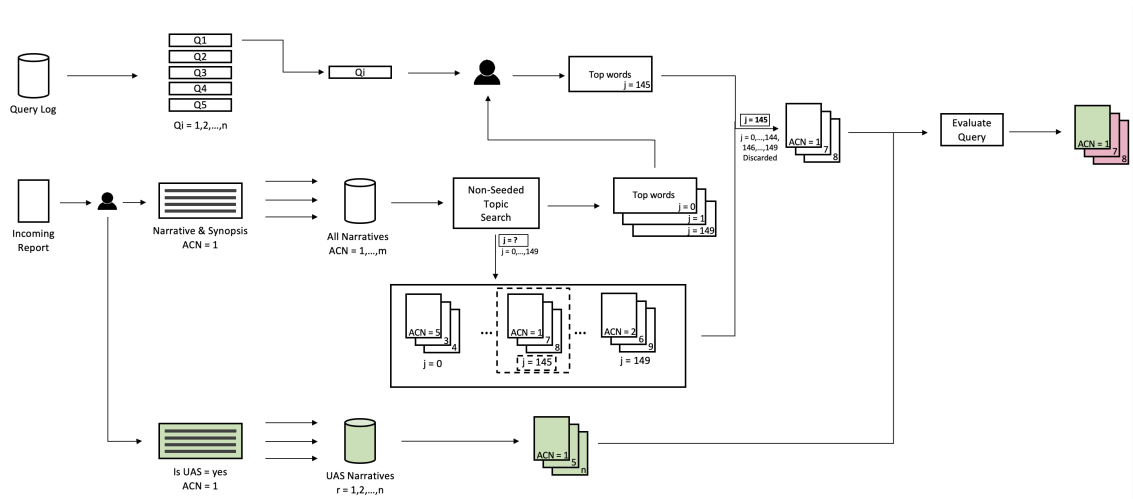
We last consider how classic topic model could be used in the context of a query, i.e. the experimental setup in Figure 3 used in RQ1. Here, because the topic model does not accept words, the query Q_i can't be used. Nonetheless, the user querying the system is assumed to still be interested in UAS reports. Since seed words can't be used, we lose the guarantee topic $j = 0$ is related to the query, and now the user must decide which topic j should be examined for UAS reports. Because topic modeling generate a set of words for each topic (commonly ten words are displayed), this set of words can be used to decide on which topic j should be examined. Since we chose $j = 150$ as a consequence of model tuning, this thus require the user to browser through 150 sets of 10 words, or 1500 words total.

In our discussion of results, we simulated this step, by looking for UAS related words among the 150 set of words, which led to the choice of topic j = 145 depicted in the diagram of Figure 3, and which we discuss in the following results section. We can see more clearly here the contribution of the method proposed in this work: By using seed topic modeling, no overheard is required on the user part to manually browse through a large number of set of words in order to identify the topic of interest, to then decide which group of reports should be browsed. While this approach is less time consuming than exhaustively look through all reports directly, which would not be viable, it is still not a reasonable approach for the purposes of a query system.

IV. Results

RQ1 What is the performance of an UAS query using classic topic model?

Fig. 3 Experiment Setup for RQ1. What is the UAS reports discoverability?



To answer RQ1, we used the experimental setup of Figure 3. We chose the manually identified topic j = 145, as its set of words contained UAS related terms that we deem general enough. These were:
Topic j = 145 top 10 words: **mission, link, remote, program, 107, function, DJI, automatically, reported, creates**
.

To simulate a user browsing through the associated set of reports of topic j=145, with the intent to discover UAS reports, we then evaluated the performance of the reports assigned to this topic, when compared to our ground truth (i.e. the list of all known UAS reports at the given time period). The confusion matrix in table 4 shows the associated results.

Table 4 Performance of classic topic model repurposed for query.

		Reference		
		No	Yes	Total
Prediction	No	2137	56	2193
	Yes	172	26	198
	Total	2309	82	2391

From the confusion matrix, we can see a total of 2391 reports were available in the system, out of which j = 0 to j = 149 topics were created, partitioning these reports. By choosing topic j = 145 for containing UAS related terms, a total of 172 + 26 = 198 reports were assigned to topic j = 145, and therefore were retrieved to the user. Out of the 198 retrieved reports, only 26 of these reports were UAS related. Note while the topic model is unaware of which reports are UAS related, we are still able to evaluate the results because of the UAS labels provided by the expert analysts. We can also observe from the confusion matrix that 56 UAS reports were not assigned to j = 145. Overall, we conclude the results of using an entirely unsupervised approach are less than desirable for the user, both for its overhead in identifying a topic from the set of words, but also in the number of UAS reports which would be incorrectly provided and missed.

RQ2 What is the performance of each UAS query by the DBOL and proposed model system?

To answer RQ2, we use the experimental setup in Figure 1. Because now the query words Q_i are actually used in the system, we can compare the DBOL Search against the Seed Topic Search for each of the five queries in table 1 highlighted in bold. Moreover, because topic j = 0 is guaranteed to be related to the list of seed terms, this approach does not require any user overhead to be provided with reports.

Observe also as we move from query 1 through 5 the number of terms in the query tends to increase, however since the order of queries reflects their usage frequency, shorter queries were preferred by users during this time period for general UAS queries. Table 5 summarizes the precision and recall for both methods across all five queries, and trade-offs between precision and recall are emphasized in bold between both methods. For the "UAV" query, both methods under-perform. In the following subsection we discuss in more detail these results using the confusion matrices.

Table 5 Precision and Recall for the five selected queries

Query	DBOL Precision	Topic Precision	DBOL Recall	Topic Recall
(DRONE)	**0.98**	0.69	0.74	**0.95**
(UAV)	0.93	0.13	0.17	0.18
(UAS)	**0.98**	0.62	0.68	**0.87**
(DRONE OR UAS OR UAV)	**0.96**	0.75	**0.96**	0.97
(UAV OR UAS OR UNMANNED OR DRONE)	**0.95**	0.21	**0.96**	0.97

A. Query 1: (DRONE)

Topic $j = 0$ top 10 words: **drone, uas, reported, uav, dji, object, drones, authorization, 107, color**

.

The confusion matrices in Table 6 compare the performance of our first query using the term "DRONE" between the DBOL Search and and the Seed Topic Search. From the very start, we can see a substantial improvement using seed terms when compared to the confusion matrix in Table 4 in regards to the number of correctly retrieved reports (61 reports for DBOL and 78 reports for Seed Topic Research, versus only 26 from the classic topic model).

In regards to the DBOL versus the Seed Topic Search for this query, the seed topic search recovers a larger number of reports (78) when compared to the DBOL search (61), however, seed topic search incorrectly retrieves more non UAS reports (34) when compared to the DBOL (1). The list of suggested words is also helpful to the user for future queries.

Table 6 Drone Query Performance

(a) DBOL Search

		Reference		
		No	Yes	Total
Prediction	No	2308	21	2329
	Yes	1	61	62
	Total	2309	82	2391

(b) Seed Topic Search

		Reference		
		No	Yes	Total
Prediction	No	2275	4	2279
	Yes	34	78	112
	Total	2309	82	2391

B. Query 2: (UAV)

Topic $j = 0$ top 10 words: **uav, lessons, skills, fighter, knee, guessing, thanked, surprise, interestingly, circumstance**

.

For our second most used query in Table 7, UAV, we observed a large contrast to the retrieval of reports and precision for approaches. Indeed, the overall performance is worse than using classic topic modeling for search, when compared to the previous confusion matrix Table 4. Query 2 emphasizes the importance of the choice of words for both models, as the listed words by the Seed Topic Search also do not provide the user with words to improve the search.

C. Query 3: (UAS)

Topic $j = 0$ top 10 words: **uas, drone, reported, uav, dji, drones, object, authorization, enforcement, law**

.

For our third and last single term query, we observe a similar result profile as with Query 1 in table 6. Specifically, seed topic search is able to recover a larger number of UAS reports (72) than the DBOL search (56), however at the cost of a larger number of incorrect reports (44 incorrect versus only 1 from DBOL). The list of suggested words is also helpful to the user for future queries.

Table 7 UAV Query Performance

(a) DBOL Search

		Reference		
		No	Yes	Total
Prediction	No	2308	68	2376
	Yes	1	14	15
	Total	2309	82	2391

(b) Seed Topic Search

		Reference		
		No	Yes	Total
Prediction	No	2214	67	2281
	Yes	95	15	110
	Total	2309	82	2391

Table 8 UAS Query Performance

(a) DBOL Search

		Reference		
		No	Yes	Total
Prediction	No	2308	26	2334
	Yes	1	56	57
	Total	2309	82	2391

(b) Seed Topic Search

		Reference		
		No	Yes	Total
Prediction	No	2265	10	2275
	Yes	44	72	116
	Total	2309	82	2391

D. Query 4: (DRONE OR UAS OR UAV)

Topic $j = 0$ top 10 words: **drone, uas, uav, reported, dji, drones, object, authorization, 107, black**

.

In Query 4, as more terms are added to the query, we observe a different performance profile. Both DBOL and Seed Topic Search performance improve when compared to the previous queries, nearly recovering all UAS reports. Although the seed topic research number of incorrectly retrieved UAS reports decreases, it is still sufficiently inferior to the DBOL search. In this case, a trade-off is no longer observed as in past queries, and the use of the DBOL search would be preferred. Nonetheless, the list of suggested words by the seed topic research is still helpful.

Table 9 Drone OR UAS or UAV Query Performance

(a) DBOL Search

		Reference		
		No	Yes	Total
Prediction	No	2306	3	2309
	Yes	3	79	82
	Total	2309	82	2391

(b) Seed Topic Search

		Reference		
		No	Yes	Total
Prediction	No	2283	2	2285
	Yes	26	80	106
	Total	2309	82	2391

E. Query 5: (UAV OR UAS OR UNMANNED OR DRONE)

Topic $j = 0$ top 10 words: **drone, uas, uav, unmanned, reported, object, black, dji, bird, drones**

.

For our fifth and least used query of the five in the time period, we observed a similar performance profile to query 4 in regards to the number of total UAS reports recovered. However, we can notice that the number of incorrectly UAS retrieved reports substantially increases by the inclusion of word unmanned for the Seed Topic Search. Once more, the set of words proves helpful to increase the user vocabulary for future searches.

RQ3 Can the suggested list of terms of the topic model be leveraged as term suggestions?

Table 10 UAV or UAS or Unmanned or Drone Query Performance

(a) DBOL Search

		Reference		
		No	Yes	Total
Prediction	No	2305	3	2308
	Yes	4	79	83
	Total	2309	82	2391

(b) Seed Topic Search

		Reference		
		No	Yes	Total
Prediction	No	2011	2	2013
	Yes	298	80	378
	Total	2309	82	2391

Having observed that in all but the UAV query 2 the set of words provided were meaningful in assisting the user in future queries, we now consider our third research question. We observed that the answer to our RQ3 could be derived from our discussed model performances. Consider for instance the most frequent performed query, "DRONE", and the associated set of words suggested by seed topic modeling:

Query 1 Topic $j = 0$ top 10 words: **drone**, **uas**, reported, **uav**, dji, object, drones, authorization, 107, color

.

From this set of words, a user could perform a new query using the DBOL search, including the words "uas" or "uav", resulting a new query "drone or uas or uav". Note, however, this would be exactly Q4. Since Q4 performance is superior to Q1, we posit a system in which seed topic search suggests terms for users to execute on the existing DBOL Search, could improve the discoverability of UAS reports by combining both approaches.

V. Conclusion and Future Work

In this work, we proposed an alternative report retrieval system to the existing DBOL Search which is currently in operation on the ASRS. By leveraging the query log to identify representative use cases from user queries, and existing report review processes in ASRS, we devised two experimental setups to compare and build upon our prior work in improving the discoverability of ASRS reports. We observed for single word queries, a trade-off occurs between recovery and recall on the existing search and the proposed seed topic search method, but the existing DBOL search is preferred for queries with more words. By combining the summary words of seed topic search to perform queries in DBOL, however, we observed the best outcome. In future work, we hope to re-evaluate the choice of the number of topics for the models, vocabulary drifting, and expand the qualitative analysis of results. We also hope to test our experimental setups with different algorithms, such as BERTopic [14] as part of more complex pipelines [15].

Acknowledgments

This research was conducted at NASA Ames Research Center. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government.

References

[1] Paradis, C., Kazman, R., Davies, M., and Hooey, B., "Augmenting Topic Finding in the NASA Aviation Safety Reporting System using Topic Modeling," 2021. https://doi.org/10.2514/6.2021-1981, URL https://arc.aiaa.org/doi/abs/10.2514/6.2021-1981.

[2] Paradis, C. V., Kazman, R., Davies, M. D., and Hooey, B. L., A Survey Protocol to Assess Meaningfulness and Usefulness of Automated Topic Finding in the NASA Aviation Safety Reporting System, ???? https://doi.org/10.2514/6.2021-2404, URL https://arc.aiaa.org/doi/abs/10.2514/6.2021-2404.

[3] Srinivasan, P., Nagarajan, V., and Mahadevan, S., Mining and Classifying Aviation Accident Reports, ???? https://doi.org/10.2514/6.2019-2938, URL https://arc.aiaa.org/doi/abs/10.2514/6.2019-2938.

[4] Hu, X., Wu, J., and He, J., Textual Indicator Extraction from Aviation Accident Reports, ???? https://doi.org/10.2514/6.2019-2939, URL https://arc.aiaa.org/doi/abs/10.2514/6.2019-2939.

[5] Zhang, X., and Mahadevan, S., A Hybrid Data-Driven Approach to Analyze Aviation Incident Reports, ???? https://doi.org/10.2514/6.2018-3982, URL https://arc.aiaa.org/doi/abs/10.2514/6.2018-3982.

[6] Tanguy, L., Tulechki, N., Urieli, A., Hermann, E., and Raynal, C., "Natural language processing for aviation safety reports: From classification to interactive analysis," Computers in Industry, Vol. 78, 2016, pp. 80–95. https://doi.org/https://doi.org/10.1016/j.compind.2015.09.005, URL https://www.sciencedirect.com/science/article/pii/S0166361515300464, natural Language Processing and Text Analytics in Industry.

[7] Luo, Y., and Shi, H., "Using lda2vec Topic Modeling to Identify Latent Topics in Aviation Safety Reports," 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), 2019, pp. 518–523. https://doi.org/10.1109/ICIS46139.2019.8940271.

[8] Oza, N., Castle, J. P., and Stutz, J., "Classification of Aeronautics System Health and Safety Documents," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Vol. 39, No. 6, 2009, pp. 670–680. https://doi.org/10.1109/TSMCC.2009.2020788.

[9] Switzer, J., Khan, L., and Muhaya, F. B., "Subjectivity classification and analysis of the ASRS corpus," 2011 IEEE International Conference on Information Reuse & Integration, 2011, pp. 160–165. https://doi.org/10.1109/IRI.2011.6009539.

[10] Paradis, C., Kazman, R., Davies, M. D., and Hooey, B. L., "Identifying Emerging Safety Threats through Topic Modeling in the Aviation Safety Reporting System: A COVID-19 Study," 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), 2021, pp. 1–8. https://doi.org/10.1109/DASC52595.2021.9594369.

[11] Chen, J., Li, K., Zhu, J., and Chen, W., "WarpLDA: a Simple and Efficient O(1) Algorithm for Latent Dirichlet Allocation." CoRR, Vol. abs/1510.08628, 2015. URL http://dblp.uni-trier.de/db/journals/corr/corr1510.html#0001LZC15.

[12] Lu, B., Ott, M., Cardie, C., and Tsou, B. K., "Multi-Aspect Sentiment Analysis with Topic Models," Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, IEEE Computer Society, USA, 2011, p. 81–88.

[13] Gallagher, R., Reing, K., Kale, D., and Steeg, G. V., "Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge," Transactions of the Association for Computational Linguistics, Vol. 5, No. 0, 2017, pp. 529–542. URL https://transacl.org/ojs/index.php/tacl/article/view/1244.

[14] Grootendorst, M., "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," , 2022. https://doi.org/10.48550/ARXIV.2203.05794, URL https://arxiv.org/abs/2203.05794.

[15] Andrade, S., and Walsh, H., "Machine Learning Framework for Hazard Extraction and Analysis of Trends (Heat) in Wildfire Response," SSRN, 2022. https://doi.org/http://dx.doi.org/10.2139/ssrn.4218602.