ERDC/ITL SR-21-10

Information Technology Laboratory

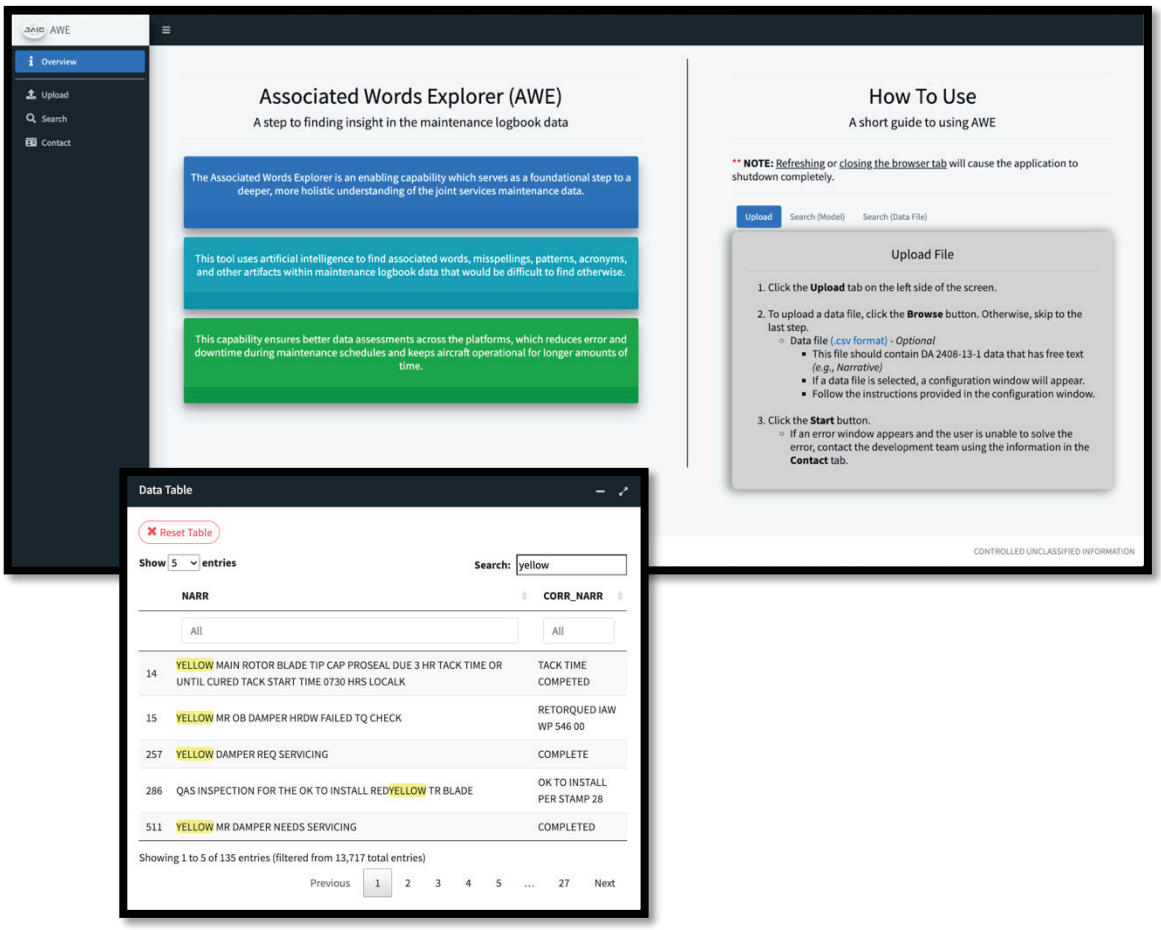
US Army Corps of Engineers®
Engineer Research and Development Center

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Associated Words Explorer (AWE) User Manual

Joshua Church, LaKenya Walker, and Amy Bednar

September 2021



The U.S. Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdc.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at https://erdclibrary.on.worldcat.org/discovery.

ERDC/ITL SR-21-10
September 2021

Associated Words Explorer (AWE) User Manual

Joshua Church, LaKenya Walker, and Amy Bednar

Information Technology Laboratory
US Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Final Report

Prepared for Joint Artificial Intelligence Center (JAIC)
Defense Information Systems Agency DISA General Fund Division
6910 Cooper Ave.
Fort Meade, MD 20755

Under MIPR HC1085015834

Abstract

This manual is intended for new users with minimal or no experience with using the Associated Word Explorer (AWE) tool. The goal of this document is to give an overview of the main functions of AWE. The primary focus of this document is to demonstrate functionality.

Every effort has been made to ensure this document is an accurate representation of the functionality of the AWE tool. For additional information about this manual, contact ERDC.JAIC@erdc.dren.mil.

Contents

Figures

Preface

This study was conducted for the Joint Artificial Intelligence Center (JAIC) under MIPR HC1085015834 under Project "ERDC PMx Product Transition to the Joint Artificial Intelligence Center (JAIC) Joint Common Foundation (JCF)." The technical monitor was Dr. Maria Niki Goerger.

The work was performed by the Computational Analysis Branch of the Computational Science and Engineering Division, U.S. Army Engineer Research and Development Center, Information Technology Laboratory (ERDC-ITL).

At the time of publication, Mr. Joshua Church was Acting Branch Chief; Dr. Jeffrey Hensley was Division Chief; and Dr. Robert Wallace was Technical Director for the Engineered Resilient Systems (ERS) program. The Deputy Director of ERDC-ITL was Dr. Jackie Pettway, and the Director was Dr. David Horner.

COL Teresa A. Schlosser was the Commander of ERDC, and Dr. David W. Pittman was the Director.

1 Introduction

1.1 Background

Across the Department of Defense (DoD), the method for performing vehicle maintenance varies between the Joint services. In an effort to better understand this process, the Joint Logistics National Mission Initiative (NMI) was created under the auspice of the Joint Artificial Intelligence Center (JAIC) to (1) improve military vehicle readiness and (2) reduce costs for unscheduled vehicle maintenance. Each maintenance action is conducted by a maintainer who logs specific details of the vehicle. One of the fundamental challenges is establishing a commonality in vernacular when describing the maintenance problem (the narrative) and the action performed to mitigate or fix the problem (the corrective action). To address this issue, the Associated Words Explorer (AWE) tool was created to analyze the maintenance logbook data.

1.2 Purpose

AWE is an enabling capability which serves as a foundational step to a deeper, more holistic understanding of the Joint services maintenance data. This application uses natural language processing (NLP) techniques to find associated words, misspellings, patterns, acronyms, and other artifacts within maintenance logbook data that would be difficult to find otherwise. This capability enables better data assessments across the platforms, which will reduce error and downtime during maintenance schedules and keep vehicles operational for longer periods of time.

1.3 Users

This section discusses the different types of user.

1.3.1 Maintainer

A maintainer should utilize AWE to to understand the commonality in vernacular among the different maintainers. For example, one maintainer may use the word “bld” as a shorthand version of “blade.” AWE allows maintainers to find these patterns to better understand descriptions of maintenance events.

1.3.2 Analyst

An analyst should utilize the output from AWE to find patterns within the free-text fields of the logbook data. An analyst may need guidance from a subject matter expert in selecting key words to choose.

1.4 Requirements

This section covers the following requirements to use AWE: *operating system*, *hardware*, *software*, and *data requirements*.

1.4.1 Recommended operating systems

- *Windows:* Windows 10
- *Mac:* macOS Sierra or newer
- *Linux:* Ubuntu.

1.4.2 Hardware

It is recommended to have a computer less than five years old that has the following:

- Processor: minimum 2 GHz; recommended 3 GHz
- Hard Drive: minimum 32 GB; recommended 64 GB
- Memory (RAM): minimum 4 GB; recommended 8 GB.

1.4.3 Software

Supported Browsers are

- Google Chrome, and
- Firefox.

Other important software includes

- R;
- RStudio; and
- Java 8, 9, 10, 11, 12, or 13.

1.4.4 Data

AWE is designed to work with Department of the Army (DA) 2408-13-1 maintenance logbook data. The DA 2408-13-1 form contains aircraft inspection and maintenance records. The data in these records include the following:

- Operational remarks or faults found during flight or operation, pre-flight, through flight, and post flight inspections
- Checks, services, and scheduled or unscheduled maintenance inspections
- Faults found including battle damage assessment and repair, when faults were found, when faults occurred, how faults were recognized, effect faults had on the operation or mission, and corrective actions taken to correct faults
- Work-hours it takes to do maintenance and quality control work by military, civilian, and contract maintenance support personnel
- Condition status symbols for aircraft and mission-related equipment.

NOTE: This data file must be in comma-separated variables (.csv) format.

1.4.5 Columns (Logbook Fields)

The following columns are generally used for AWE:

- **NARR**: 2408-13-1 write-up narrative
- **CORR_NARR**: 2408-13-1 write-up of corrective narrative.

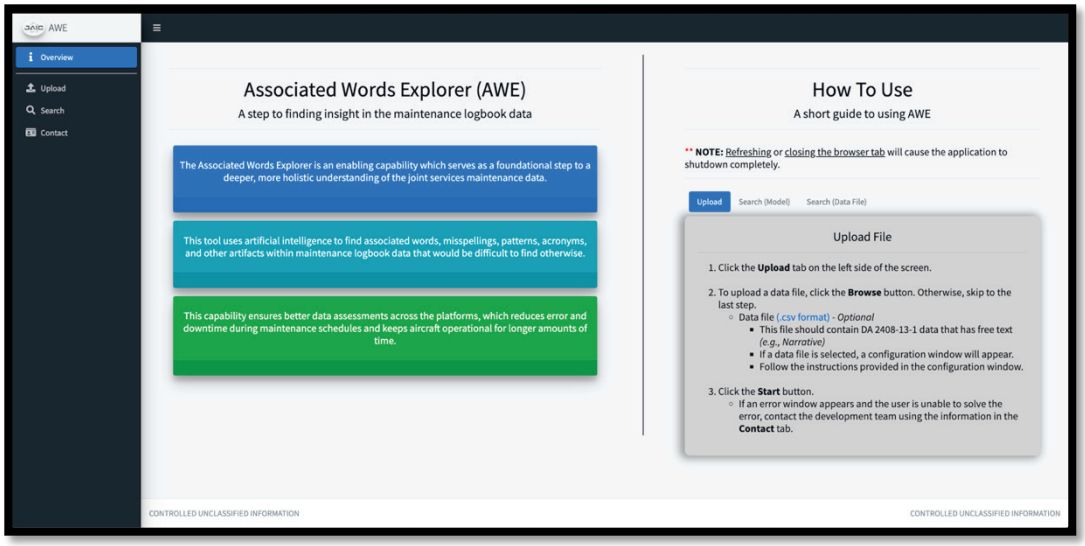
NOTE: The user can select which columns to search via the AWE tool.

2 User Interface

2.1 Main page

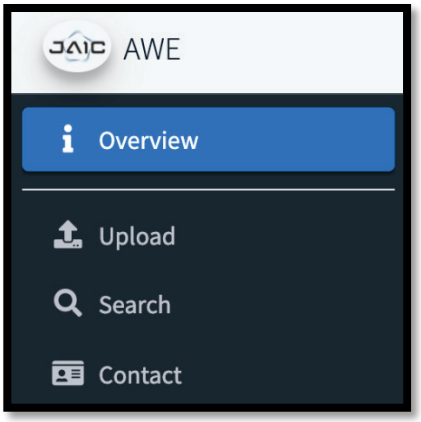
Once the AWE tool loads, the user is presented with the interface shown in Figure 1.

Figure 1. Main page.



On the left side of the screen, the user can select different tab options. There are four tabs available for the user to select: **Overview**, **Upload**, **Search**, and **Contact** (see Figure 2). Information about each tab can be found in the following sections.

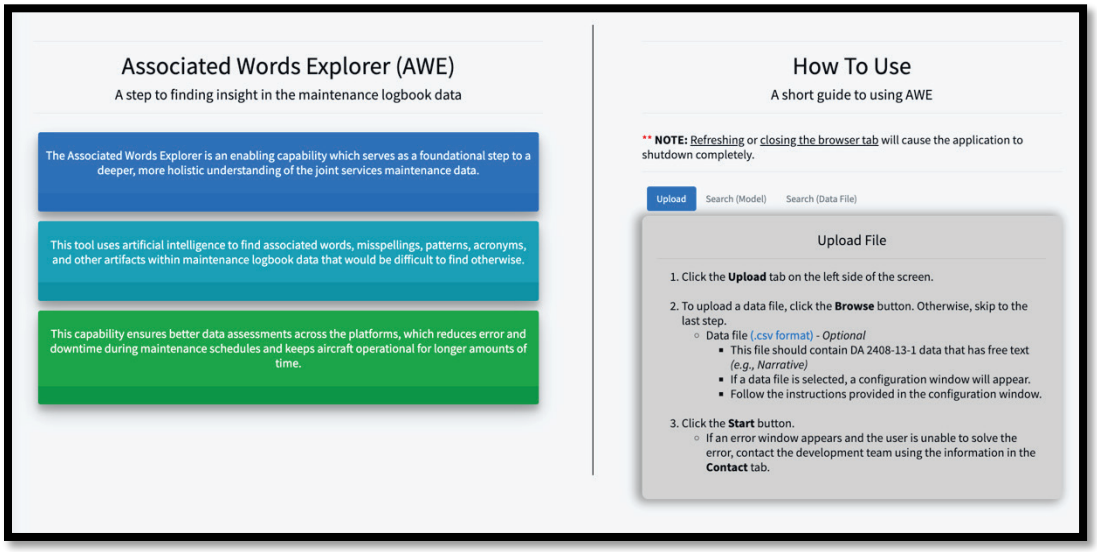
Figure 2. Tabs detail.



2.2 Overview

The **Overview** tab contains information on the purpose of AWE and brief instructions on how to use the software (see Figure 3). These instructions are designated for the **Upload** and **Search** sections.

Figure 3. Overview tab.

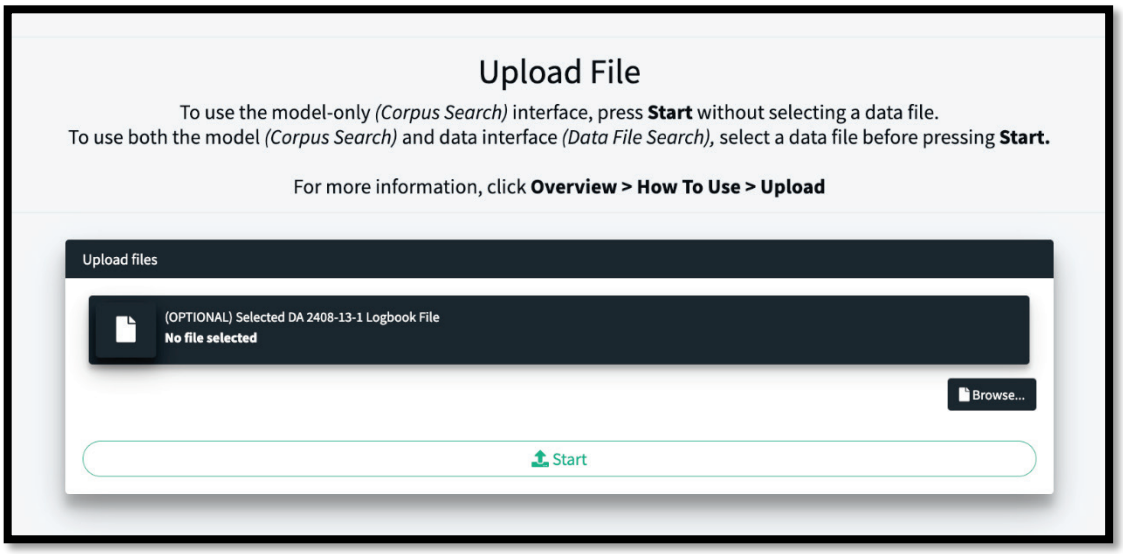


NOTE: When using the AWE tool, refreshing or closing the browser tab will cause the application to shut down.

2.3 Upload

The **Upload** tab allows the user to upload a data file. After clicking the **Upload** tab, the user is presented with an option to upload a DA 2408-13-1 logbook data file (see Figure 4).

Figure 4. Upload window.



Uploading a data file is an optional step. The data file requirements are as follows:

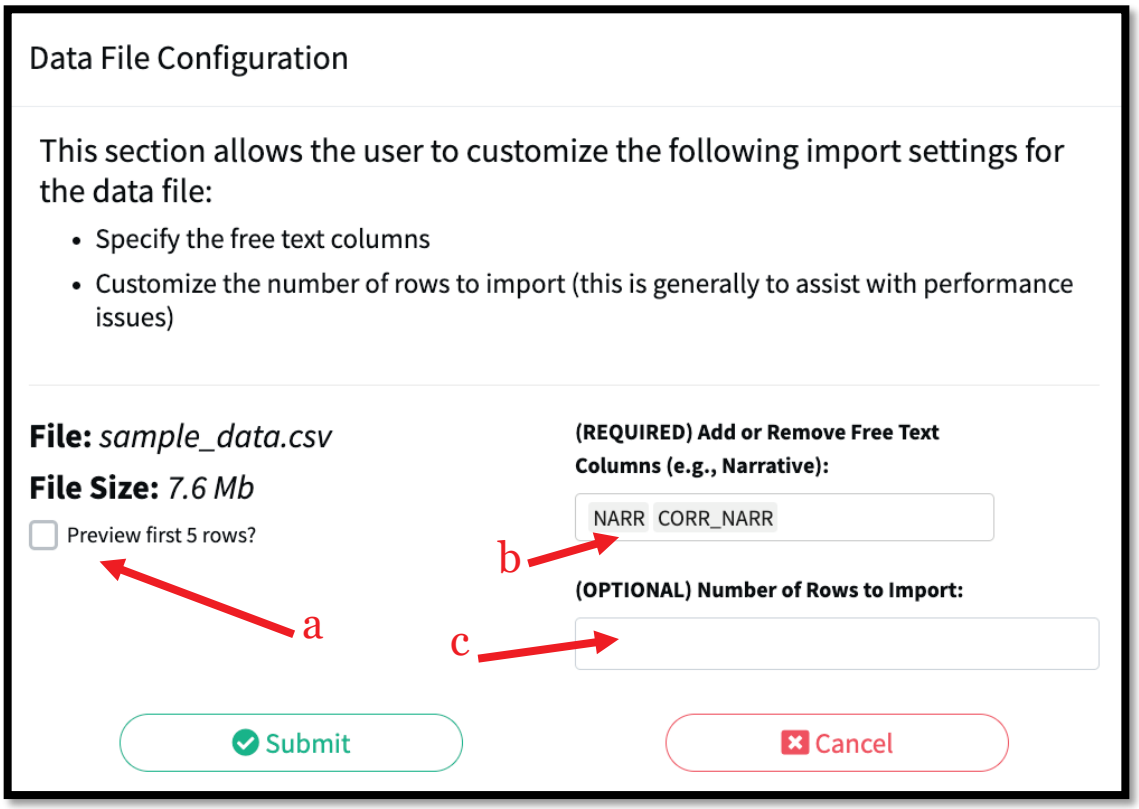
- This file should be DA 2408-13-1 data that contains free text. Free text is most commonly found in the "Narrative" and "Corrective Action" columns in the data file.
- The data file must be in a comma-separated values (.csv) format.

NOTE: This process can be skipped by pressing the *Start* button.

To upload a file, perform the following steps:

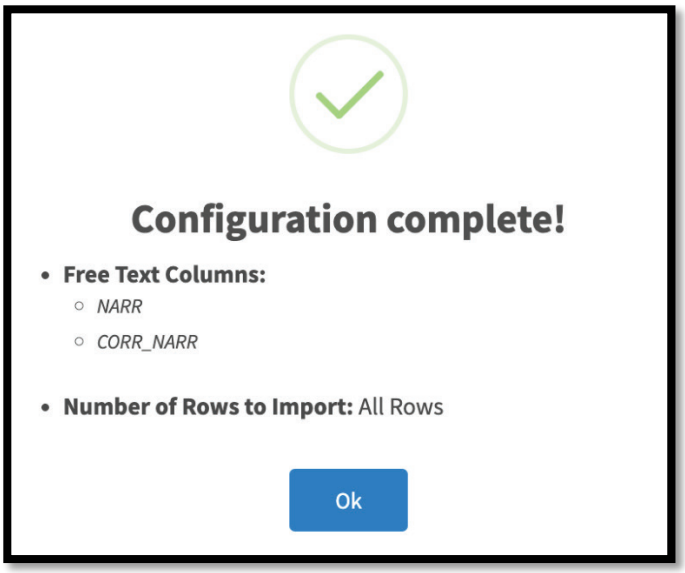
1. Click Browse
2. Navigate to the system path that contains the DA 2408-13-1 logbook file.
    a. Once selected, the "Data File Configuration" window will appear (see Figure 5).

Figure 5. Data file configuration window.



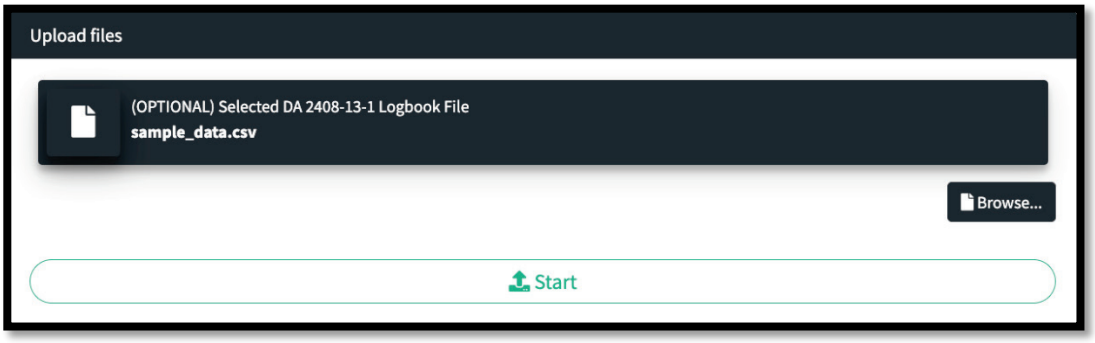
3. In the *Data File Configuration* window, the user is presented with the following options:
    a. *Preview Rows*
        i. This is an optional step. This allows the user to preview the first five rows of the selected data file.
    b. *Add or Remove Free Text Columns*
        i. This is a required step. The tool tries to determine which columns are free text columns. The user should add columns with free text or remove columns that are not free text columns.
    c. *Number of Rows to Import*
        i. This is an optional step. This allows the user to customize the number of rows to import from the data file. If none are selected, all the rows will be imported.
4. After selecting the configuration options, click *Submit*.
    a. If the user selects *Cancel*, all the columns and rows from the selected data file will be imported.
5. Click *Ok* to close the *Confirmation* window (see Figure 6).

Figure 6. Example of configuration confirmation.



6. Click *Start* (see Figure 7).

Figure 7. Uploading required files window.



a. A sequence of loading windows will appear after this operation has been performed (see Figures Figure 8, Figure 9, and Figure 10).

Figure 8. Loading screen—model.

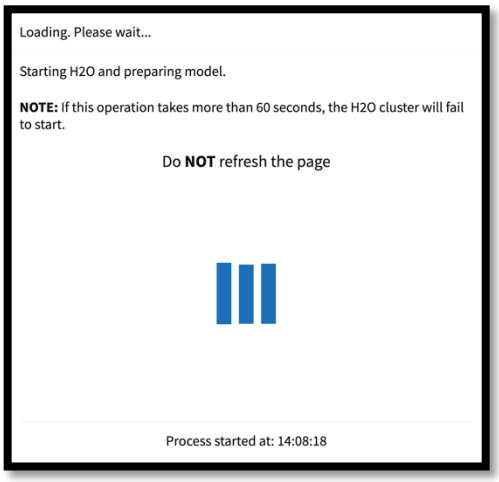
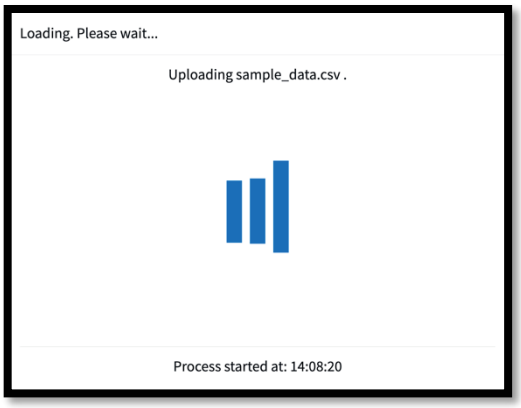
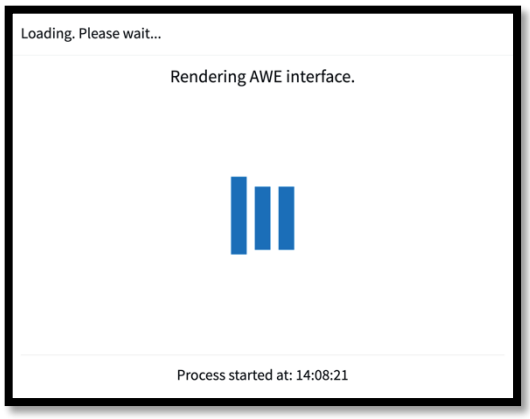


Figure 9. Loading screen—data file.



This window will only appear if a data file was uploaded.

Figure 10. Loading screen—rendering AWE.



The **Search** functionality can now be used. This is covered in the following section.

2.4 Search

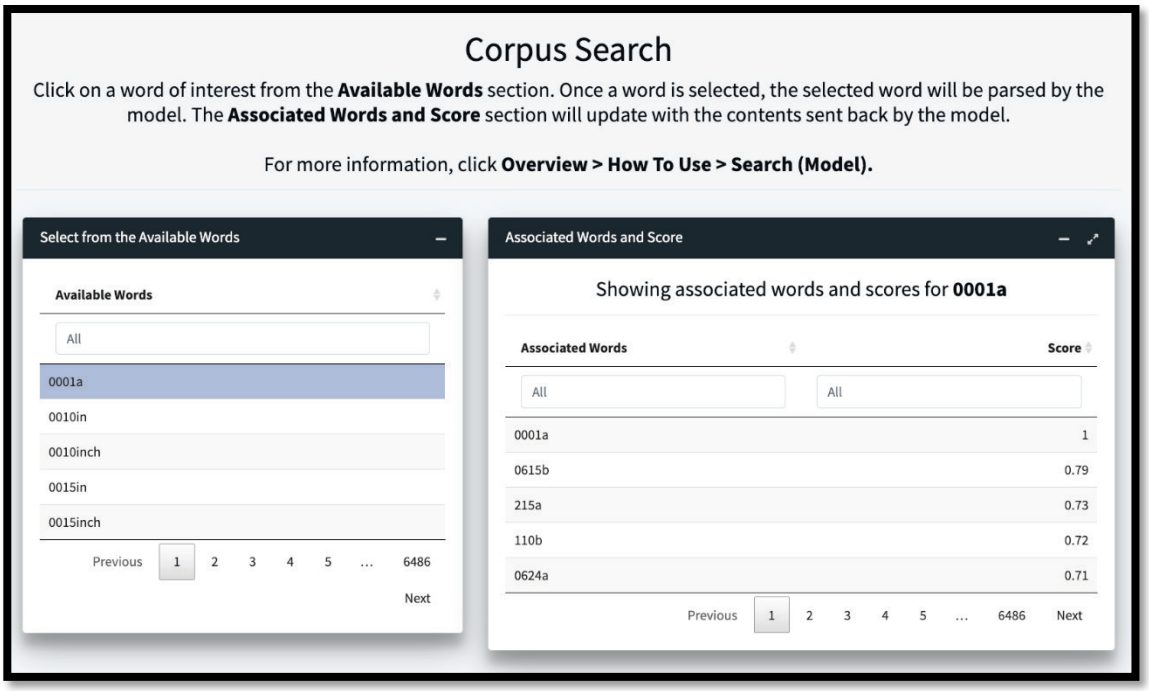
The **Search** tab allows a user to search through a natural language processing (NLP) model that has been trained on the DA 2408-13-1 logbook data. Optionally, if a data file was provided in the steps above, the user can search the data.

NOTE: It is assumed that the user has stepped through the **Upload** section above.

2.4.1 Model interface

In the *Corpus Search* section, the user can search for a word of interest, which will be parsed by the model. The respective associated words are returned in table format for the user to explore (see Figure 11).

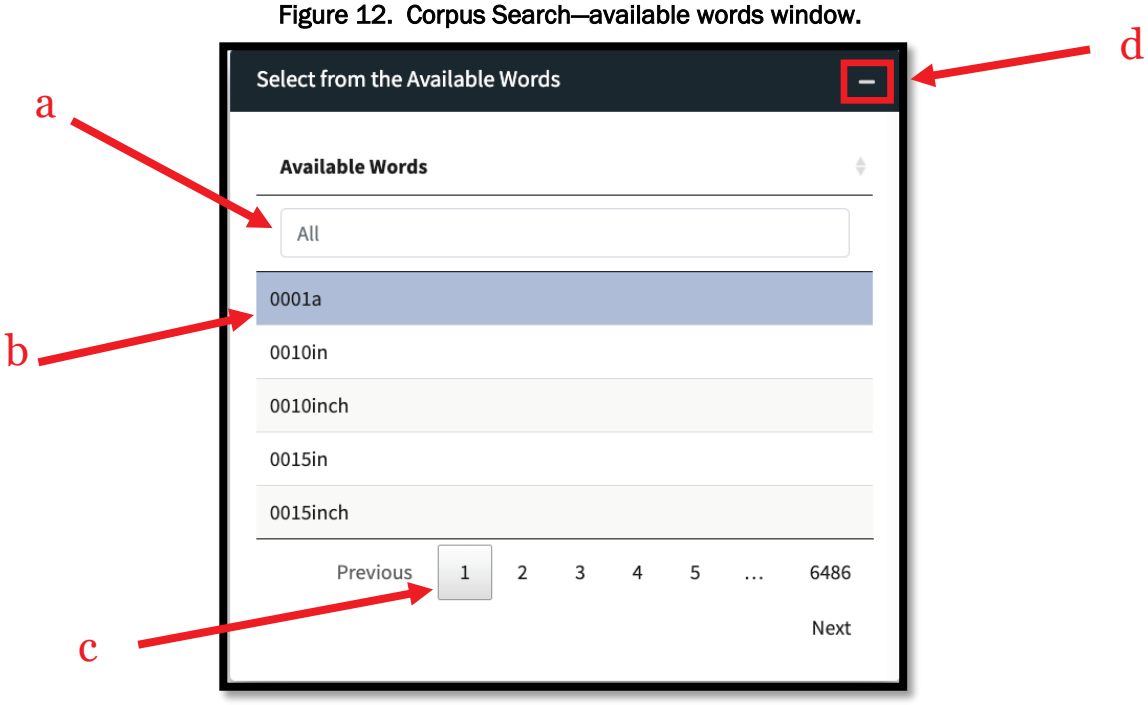
Figure 11. Model interface (Corpus Search).



2.4.1.1 Available words

This component provides a searchable table that contains the available words trained by the model (see Figure 12). There are over 32,000 unique words available for a user to select from.

Figure 12. Corpus Search—available words window.

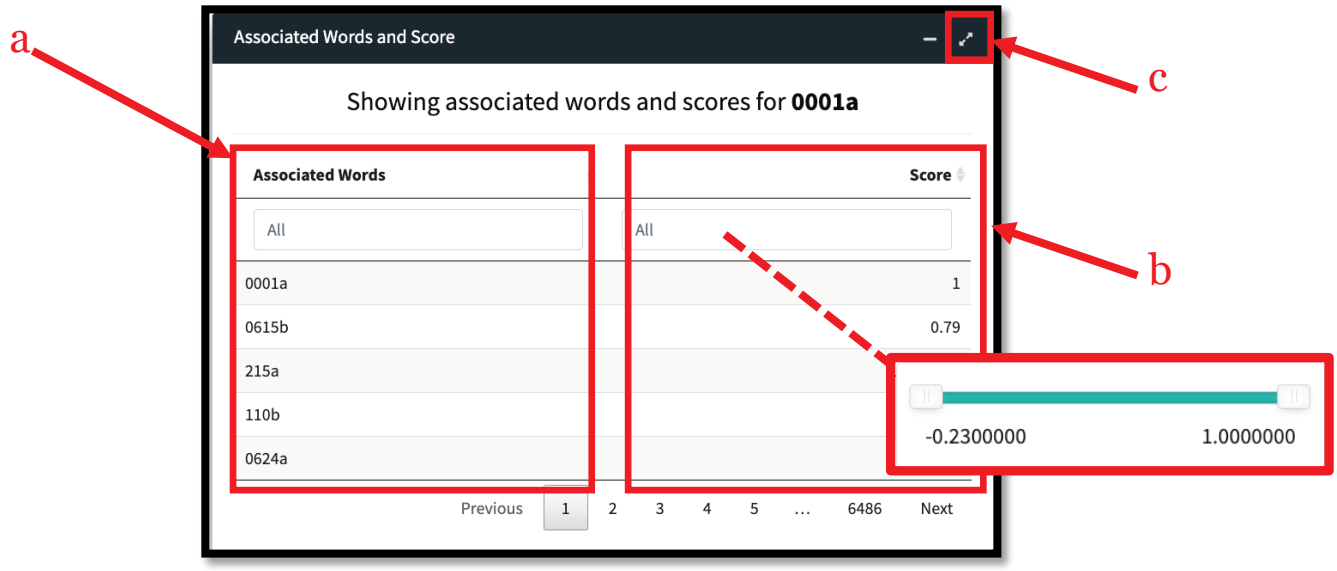


a. *Search Bar:* This feature allows the user to filter and search for a word of interest.
b. *Word Selection:* Once a word of interest has been located, the user should click the corresponding row. Clicking a word populates the **Associated Words and Score** component, which is discussed in the following section.
c. *Pages*: This numbers display the current page within the table. The user can navigate through the different pages here. This feature is also available in the **Associated Words and Score** and **Data Table** components.
d. *Minimize*: This option allows the user to minimize the **Available Words** window. This feature is also available in the **Associated Words and Score** and **Data Table** components.

2.4.1.2 Associated words and score

This component provides a table of associated words (and their respective score) in response to the user-selected word from the **Available Words** component (see Figure 13).

Figure 13. Corpus Search—associated words and score window.



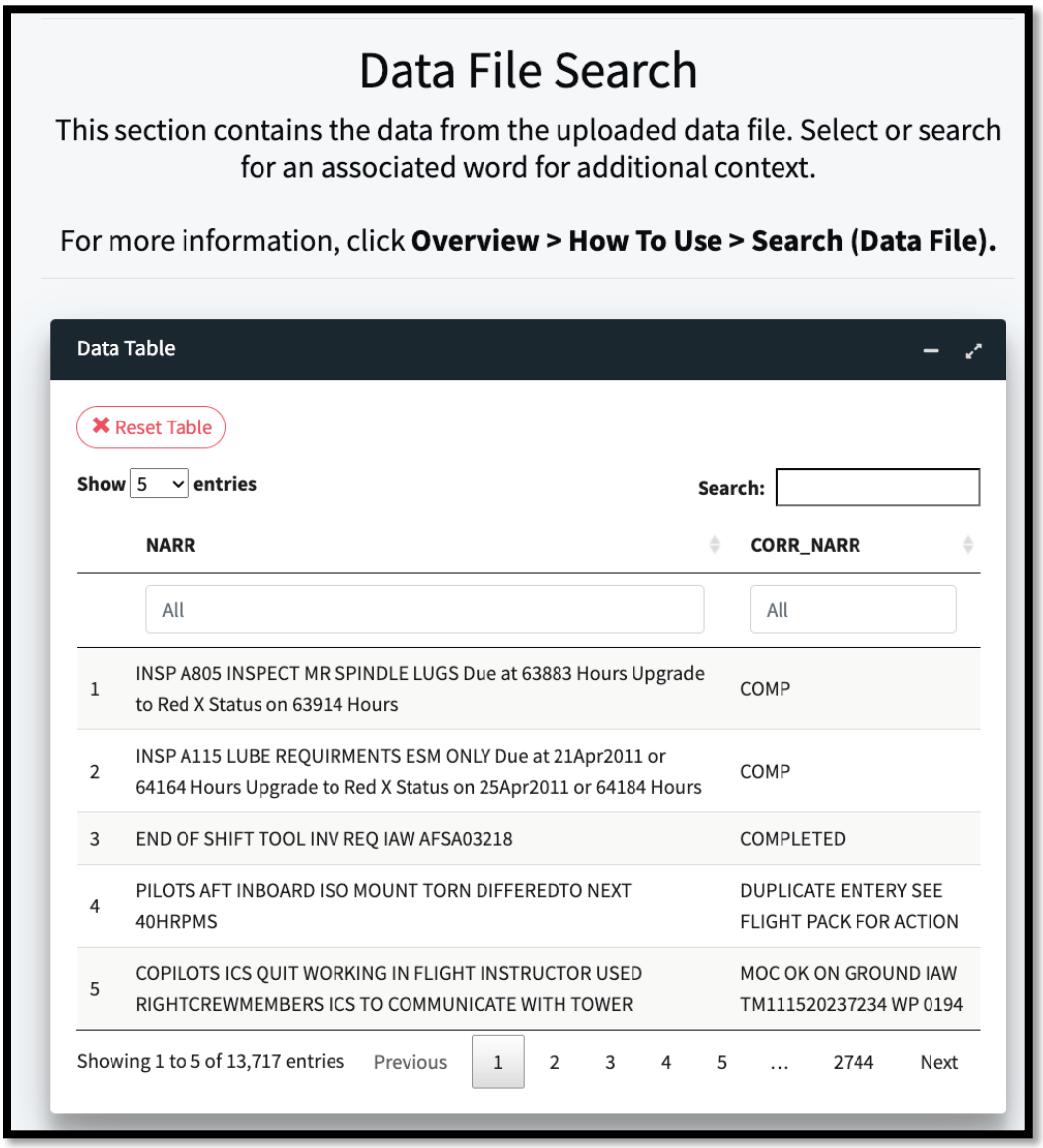
a) *Associated Words*: A list of associated words related to the user-selected word from the **Available Words** section
b) *Score*: The score, calculated by its cosine similarity, shows the relationship between the user-selected word and its respective associated words. The user can filter the score by using the available slider bar.
   - A score closer to 1 means it is a similar meaning.
   - A score of 0 means it is independent (not associated—also known as orthogonal).
   - A score closer to -1 means opposing meaning.
c) *Maximize*: This option allows the user to maximize the **Associated Words and Score** window. This feature is also available in the **Data Table** component.

2.4.2 Data interface

In the *Data File Search* section, the user can search through the provided data file (see Figure 14).

NOTE: This section is optional and will only appear if a data file was uploaded.

Figure 14. Data interface (Data File Search).



2.4.2.1 Data table

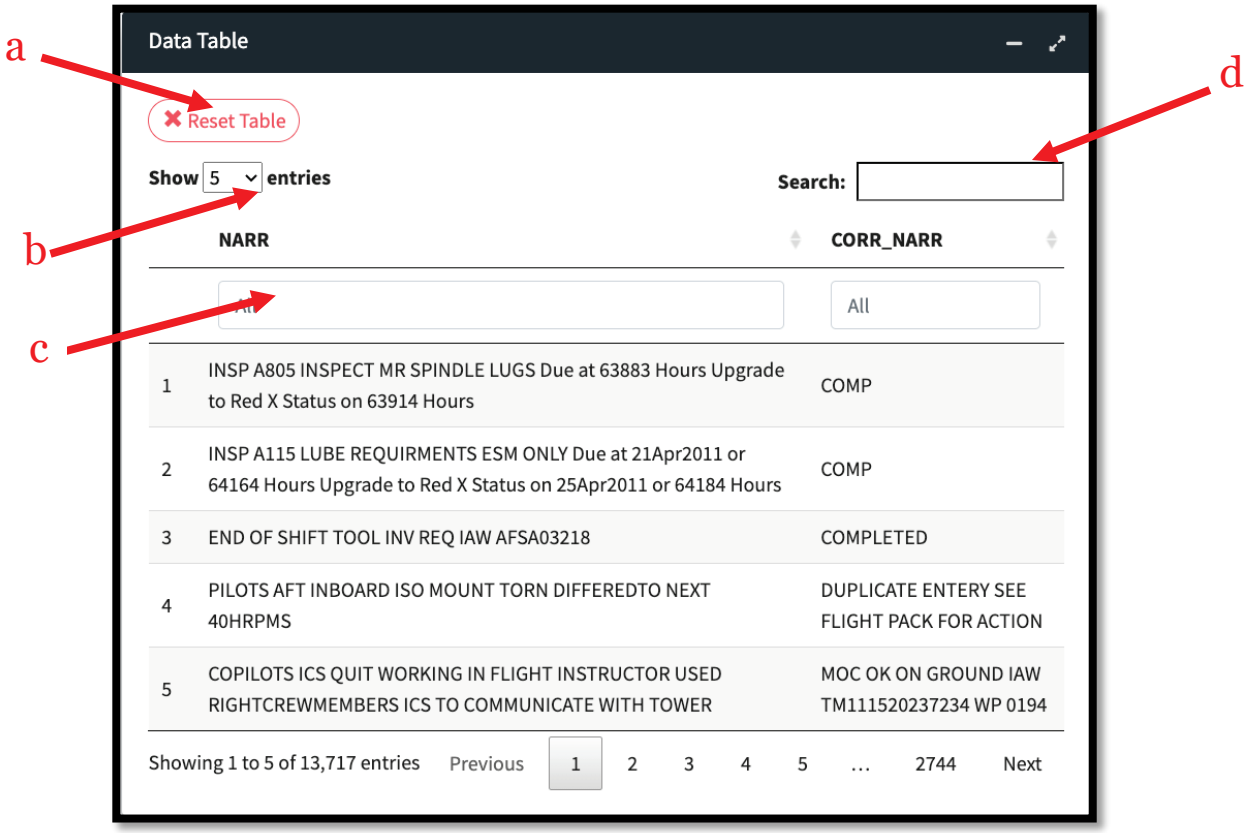
This component allows the user to search through the provided data file (see Figure 15).

a) *Reset Table*: This allows the user to reset the data table back to its original state.
b) *Row Count*: This allows the user select the number of rows shown in the window.
c) *Column Search*: This allows the user to search or filter words in the respective column.

d) *Full Search*: This allows the user to search or filter words amongst all of the columns.

- Auto *Search*: After searching for associated words from the *Corpus Search* section, the user can click on an associated word of interest to automatically search for the word in the provided data file. For an example of this feature, navigate to the **Example Walkthrough** section below.

Figure 15. Data file search—data table.

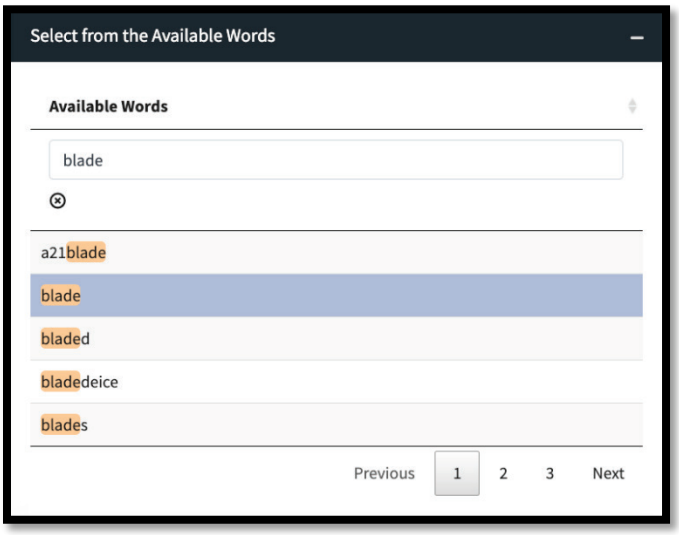


2.4.2.2 Example walkthrough

This section walks the user through the process for looking up the word "blade" and exploring its related associated words (see Figure 16). This example assumes the user has uploaded a data file.

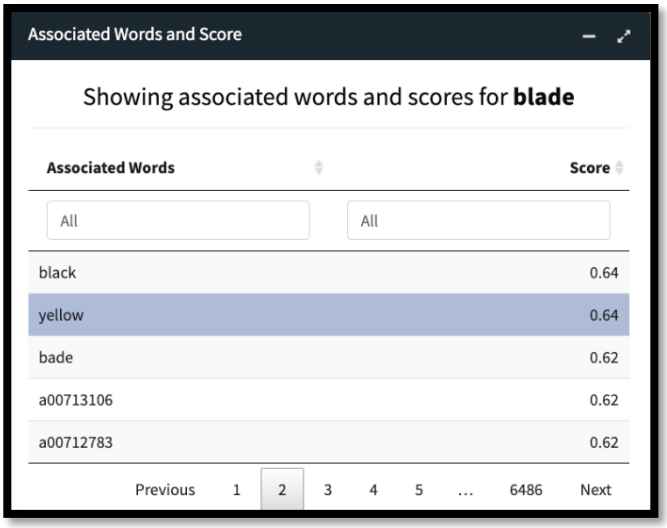
1. In the **Available Words** component, enter the word "blade" in the search bar.
2. Click the row with the word "blade."

Figure 16. Example walkthrough—search and select.



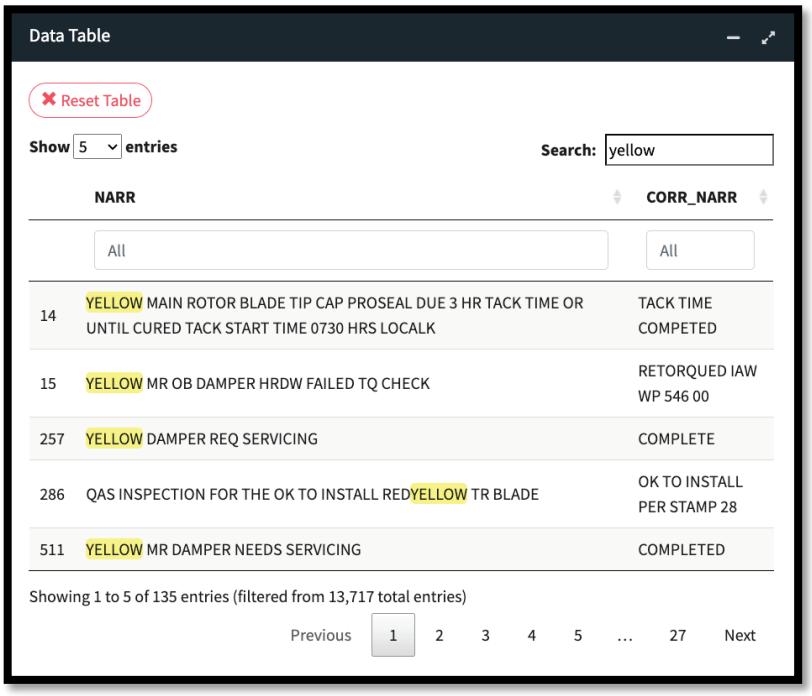
3. In the **Associated Words and Score** component, click on an associated word of interest. In this example, the word "yellow" will be used (see Figure 17).

Figure 17. Example walkthrough—word selected.



   a. After clicking "yellow," the **Data Table** component will automatically look up "yellow" to find additional context in the uploaded data (see Figure 18). If the provided data file does not contain the selected word, then the table will be empty.

Figure 18. Example walkthrough—additional context.



4. This concludes the process for finding associated words for "blade." Repeat steps 1–3 for additional words of interest.

2.5 Contact

The **Contact** tab provides the contact information for the developer of the AWE tool (see Figure 19).

Figure 19. Contact information.

3 Summary

AWE is a capability which enables better data assessments across platforms. It serves as a foundational step to a deeper, more holistic understanding of the Joint services maintenance data by using NLP techniques.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
09-2021	Summary Report	

4. TITLE AND SUBTITLE
Associated Words Explorer (AWE) User Manual

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT

6. AUTHOR(S)
Joshua Church, LaKenya Walker, and Amy Bednar

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Information Technology Laboratory
US Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/ITL SR-21-10

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Joint Artificial Intelligence Center
Defense Information Systems Agency
DISA General Fund Division
6910 Cooper Ave.
Fort Meade, MD 20755

10. SPONSOR/MONITOR'S ACRONYM(S)
JAIC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
MIPR HC1085015834

14. ABSTRACT

This manual is intended for new users with minimal or no experience with using the Associated Word Explorer (AWE) tool. The goal of this document is to give an overview of the main functions of AWE. The primary focus of this document is to demonstrate functionality.

Every effort has been made to ensure this document is an accurate representation of the functionality of the AWE tool. For additional information about this manual, contact ERDC.JAIC@erdc.dren.mil.

15. SUBJECT TERMS
Maintenance—Records—Analysis, Data processing, Natural language processing (Computer science), Machine learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (include area code)
Unclassified	Unclassified	Unclassified	SAR	25	