Volume 124, Article No. 124029 (2019) https://doi.org/10.6028/jres.124.029
Journal of Research of National Institute of Standards and Technology

Nestor: A Tool for Natural Language Annotation of Short Texts

Rachael T.B. Sexton and Michael P. Brundage

National Institute of Standards and Technology,
Gaithersburg, MD 20899, USA

rachael.sexton@nist.gov
michael.brundage@nist.gov

Software DOI: https://doi.org/10.18434/T4/1502464

Key words: maintenance work orders; manufacturing; natural language processing.

Accepted: October 30, 2019

Published: November 1, 2019

https://doi.org/10.6028/jres.124.029

1. Summary

Nestor is a software tool that annotates natural language CSV (comma-separated variable) files, with a UTF-8 (Unicode Transformation Format – 8-bit) encoding, using a process called tagging [1]. The objective of Nestor is to help analysts make their natural language data, which is often unstructured, filled with technical content, jargon, mispellings, and abbreviations, computable to improve analysis. An example of natural language data that could be input to Nestor and the subsequent output data and the corresponding output is shown in Table 1.

The annotated datasets generated by Nestor (as either a CSV or .h5 file) can be used for different analysis techniques, such as failure prediction, problem hot spot identification, and maintenance technician expertise assessment, as shown in [2–10]. Currently, the majority of use cases involve maintenance in the engineering domain (manufacturing, mining, heating ventilation and air conditioning (HVAC)), however, any natural language CSV file with UTF-8 encoding can be input to Nestor.

How to cite this article:
Sexton R, Brundage M (2019) Nestor: A Tool for Natural Language Annotation of Short Texts.
J Res Natl Inst Stan 124:124029. https://doi.org/10.6028/jres.124.029

Table 1. An example of natural language input (**Raw Text** column in this example) and subsequent output (**Item(s)**, **Problem(s)**, **Solution(s)**, **Problem(s) & Item(s)**, **Solution(s) & Item(s)** columns in this example) for Nestor. These input files often also contain other non-text based data points that can be used for other analysis, but are not directly used by Nestor.

Raw Text	Item(s)	Problem(s)	Solution(s)	Problem(s) & Item(s)	Solution(s) & Item(s)
Hyd leak at saw attachment. Replaced seal in saw attachment but still leaking - Reapirs pending with ML	Hydraulic; Saw attachment; Seal	Leak	Replaced; Repaired	Hydraulic Leak	Replaced Seal
HP Coolant pressure at 75 psi; Bad gauge/Low pressure lines cleaned ou	High Pressure Coolant; Gauge; Low Pressure Line	Broken; Low Pressure	Cleaned	Broken Gauge	Cleaned Low Pressure Line
Major hydraulic leak at SP#6 horseshoe. Repaired horseshoe seals.	Hydraulic; SP#6; Horseshoe Seal	Leak	Repaired	Hydraulic Leak	Repaired Horseshoe Seal
Clamping spool guard broken, replaced - operator could have done this!	Clamping Spool Guard; Operator	Broken	Replaced	Clamping Spool Guard Broken	N/A

2. Software Specifications

NIST Operating Unit	Engineering Laboratory, Systems Integration Division, Informational Modeling and Testing Group
Category	Analysis Graphical User Interface (GUI).
Targeted Users	Manufacturers, Maintainers, Maintenance Technicians, Analysts
Operating Systems	**Windows:** Windows 10 or greater; **Mac:** OSx v10.1 or greater; **Linux:** Linux 5.0 x86_64 or greater
Programming Language	**Executable**: *None*; **Source**: Python v3.6 or greater See https://github.com/usnistgov/nestor/tree/master/requirements
Inputs/Outputs	Input: UTF-8 encoded .csv file. Output(s): Annotated .csv file, .h5 file dashboard.
Documentation	User's Guide - https://nestor.readthedocs.io/en/latest/index.html Source Code: https://github.com/usnistgov/nestor
Disclaimer	https://www.nist.gov/disclaimer

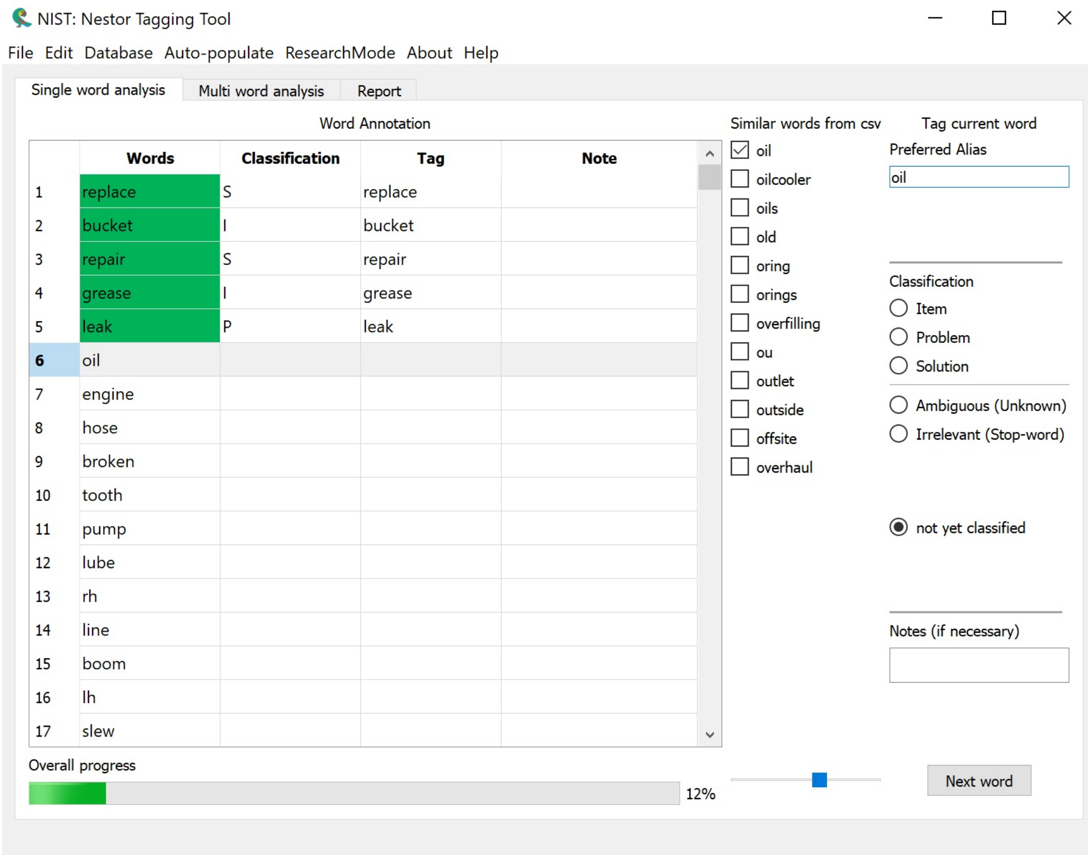

Fig. 1. A screenshot of the Nestor GUI.

3. Methods

This software provides a Graphical User Interface (GUI) (both as a standalone application[1] and the source code[2]) as seen in Fig. 1.

The software takes natural language inputs in the form of UTF-8 encoded CSV files and allows a user to select the columns containing natural language text. After columns in the CSV files are selected, the software will rank the concepts according to their frequency occurring in the data and allow the user to select similar concepts, create an alias, and provide a classification. Once the user completes this process, the software tool will automatically annotate the dataset and provide an annotated CSV and .h5 file as shown in Fig. 2. These files can then be used for various analysis techniques, such as problem identification, failure prediction, and technician skill assessment [2–7].

[1] https://www.nist.gov/services-resources/software/nestor
[2] https://github.com/usnistgov/nestor

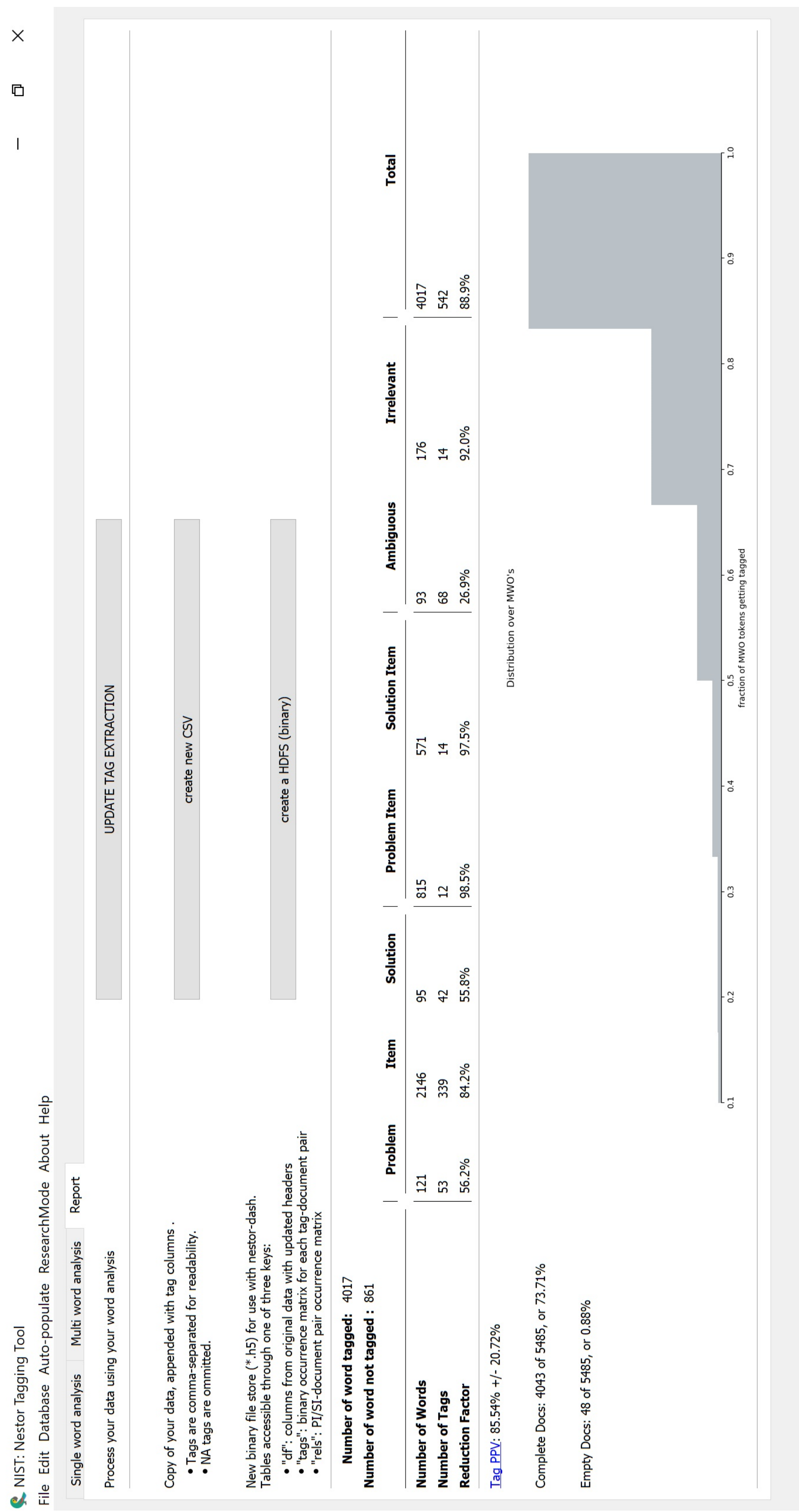

Fig. 2. A screenshot of the Nestor GUI report tab.

4. References

[1] Madhusudanan Navinchandran F, Bones L, Brundage M, Hoffman M, Moccozet S, Sexton R (2018) Nestor: a toolkit for quantifying tacit maintenance knowledge, for investigatory analysis in smart manufacturing. https://doi.org/10.18434/t4/1502464. Available at https://github.com/usnistgov/nestor

[2] Sexton R, Hodkiewicz M, Brundage MP, Smoker T (2018) Benchmarking for keyword extraction methodologies in maintenance work orders. *Proceedings of the Annual Conference of the PHM Society*, Vol. 10.

[3] Sexton R, Brundage MP, Hoffman M, Morris KC (2017) Hybrid datafication of maintenance logs from ai-assisted human tags. *2017 IEEE International Conference on Big Data (Big Data)* (IEEE), pp 1769–1777.

[4] Brundage MP, Morris K, Sexton R, Moccozet S, Hoffman M (2018) Developing maintenance key performance indicators from maintenance work order data. *ASME 2018 13th International Manufacturing Science and Engineering Conference* (American Society of Mechanical Engineers), pp V003T02A027–V003T02A027.

[5] Brundage MP, Sexton R, Hodkiewicz M, Morris KC, Arinez J, Ameri F, Ni J, Xiao G (2019) Where do we start? guidance for technology implementation in maintenance management for manufacturing. *Journal of Manufacturing Science and Engineering* 141(9):091005.

[6] Sharp M, Sexton R, Brundage MP (2017) Toward semi-autonomous information. *IFIP International Conference on Advances in Production Management Systems* (Springer, Cham), pp 425–432.

[7] Brundage MP, Kulvantunyou B, Ademujimi T, Rakshith B (2017) Smart manufacturing through a framework for a knowledge-based diagnosis system. *Proceedings of the ASME 2017 International Manufacturing Science and Engineering Conference, MSEC*, Vol. 2017, pp 1–9.

[8] Hastings E, Sexton R, Brundage MP, Hodkiewicz M (2019) Agreement behavior of isolated annotators for maintenance work-order data mining. *Proceedings of the Annual Conference of the PHM Society*, Vol. 11.

[9] Sexton R, Hodkiewicz M, Brundage MP (2019) Categorization errors for data entry in maintenance work-orders. *Proceedings of the Annual Conference of the PHM Society*, Vol. 11.

[10] Navinchandran M, Sharp ME, Brundage MP, Sexton RTB (2019) Studies to predict maintenance time duration and important factors from maintenance workorder data. *Proceedings of the Annual Conference of the PHM Society*, Vol. 11.

About the authors: *Rachael T.B. Sexton, MS is a Mechanical Engineer in the Information Modeling and Testing Group of the Systems Integration Division at NIST, currently researching the usability of natural language processing for mining useful system representations for Smart Manufacturing Systems. Their interests include statistical network analysis, Bayesian global optimization, human factors, inverse reinforcement learning, and hybrid (physics/data-driven) modeling.*

Michael P. Brundage, PhD is an Industrial Engineer in the Information Modeling and Testing Group. Dr. Brundage serves as the Project Leader for the Knowledge Extraction and Application for Manufacturing Operations project in the Model-Based Enterprise Program. Dr. Brundage's interests include Smart Manufacturing Diagnostics for Intelligent Maintenance, Sustainable Manufacturing Performance Measurement, Smart Manufacturing Capability Assessment, and Manufacturing Knowledge Visualization.

The National Institute of Standards and Technology is an agency of the U.S. Department of Commerce.