

Manuscript ID : 00001-90651

IACSE - International Journal of Cyber Security

Volume 5, Issue 1, January-June 2024, Pages 1-6, Page Count - 6



Source ID : 00000778

Implementation of a Threat Intelligence System Using Natural Language Processing on Dark Web Sources

Laura R Ashley ⁽¹⁾ Ritu M Kumar ^{(2)*} Errol T Arendz ⁽³⁾

⁽¹⁾ Threat Intelligence Analyst, United Kingdom.

⁽²⁾ Cybersecurity Data Scientist, India.

⁽³⁾ SOC Analyst, South Africa.

Abstract

The growing sophistication of cybercrime networks operating within the Dark Web has necessitated the development of automated threat intelligence (TI) systems. This paper proposes a Natural Language Processing (NLP)-based framework for extracting, analyzing, and classifying threat indicators from unstructured text within Dark Web forums, marketplaces, and hidden services. We explore various NLP techniques including Named Entity Recognition (NER), topic modeling, and sentiment analysis to assess and profile potential cyber threats. A comparative analysis of foundational methods in NLP and threat detection is included. Results highlight the potential of combining unsupervised learning with deep NLP techniques to improve threat signal extraction and reduce false positives.

Author Keywords

Threat Intelligence, Dark Web, Natural Language Processing, Cybersecurity, Topic Modeling, Entity Recognition, Threat Detection, Information Extraction, Unstructured Data Analysis, Malicious Actor Profiling, Deep Learning.

Funding Details

This research received no external funding.

ISSN Print:

Source Type: Journals

Publication Language: English

Abbreviated Journal Title: IACSE-IJCS

Publisher Name: International Academy for Computer Science and Engineering (IACSE)

Major Subject: Physical Sciences

Subject area: Information Technology and Cyber Security

ISSN Online:

Document Type: Journal Article

DOI:

Access Type: Open Access

Resource Licence: CC BY-NC

Subject Area classification: Computer Science

Source: SCOPEDATABASE

Reference

[1] Liu, Bing. *Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, 2nd ed., 2010.*

[2] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global Vectors for Word Representation." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1532–1543.*

[3] Cambria, Erik, Björn Schuller, Yunqing Xia, and Catherine Havasi. "New Avenues in Opinion Mining and Sentiment Analysis." *IEEE Intelligent Systems, vol. 28, no. 2, 2013, pp. 15–21.*

[4] Ferrucci, David, and Adam Lally. "UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment." *Natural Language Engineering*, vol. 10, no. 3-4, 2004, pp. 327–348.

[5] Piskorski, Jakub, and Roman Yangarber. "Information Extraction: Past, Present and Future." *Multi-source, Multilingual Information Extraction and Summarization*, Springer, 2012, pp. 23–49.

[6] Mohr, John W., Ronald L. Breiger, Robin Wagner-Pacifici, and Petko Bogdanov. "Graphing the Grammar of Motives in National Security Strategies: Cultural Interpretation, Automated Text Analysis and the Drama of Global Politics." *Poetics*, vol. 41, no. 6, 2013, pp. 670–700.

[7] Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoŕiuc-Pietro, and Vasileios Lampos. "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective." *PeerJ Computer Science*, vol. 2, 2016, e93.

[8] Dahl, George E., Dong Yu, Li Deng, and Alex Acero. "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition." *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, 2011, pp. 30–42.

[9] Cambria, Erik, and Björn White. "Jumping NLP Curves: A Review of Natural Language Processing Research." *IEEE Computational Intelligence Magazine*, vol. 9, no. 2, 2014, pp. 48–57.

[10] Wiegand, Michael, and Anne Schmidt. "A Survey on Hate Speech Detection Using Natural Language Processing." *Proceedings of the Workshop on Natural Language Processing for social media*, 2017, pp. 1–10.

[11] Nasukawa, Tetsuya, and Jeonghee Yi. "Sentiment Analysis: Capturing Favorability Using Natural Language Processing." *Proceedings of the 2nd International Conference on Knowledge Capture*, 2003, pp. 70–77.

[12] Chen, Hsinchun, Roger H. L. Chiang, and Veda C. Storey. "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS Quarterly*, vol. 36, no. 4, 2012, pp. 1165–1188.

[13] Young, Tom, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. "Recent Trends in Deep Learning Based Natural Language Processing." *IEEE Intelligent Systems*, vol. 33, no. 3, 2018, pp. 55–62.

[14] Savova, Guergana K., James J. Masanz, Philip V. Ogren, et al. "Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, Component Evaluation and Applications." *Journal of the American Medical Informatics Association*, vol. 17, no. 5, 2010, pp. 507–513.