CIRI | CRITICAL INFRASTRUCTURE RESILIENCE INSTITUTE

A DEPARTMENT OF HOMELAND SECURITY CENTER OF EXCELLENCE

Evaluating Ontologies for Cybersecurity Workforce

Author: Ana Robinson
Faculty Advisor: Dr. Dipakkumar P. Pravin

University of North Texas

Homeland Security Challenge

The National Scale Workforce Development Initiatives aim to enhance the Nation's cybersecurity by developing a standardized process. By understanding this process the urgent workforce needs can be addressed for the betterment of the homeland security enterprise.

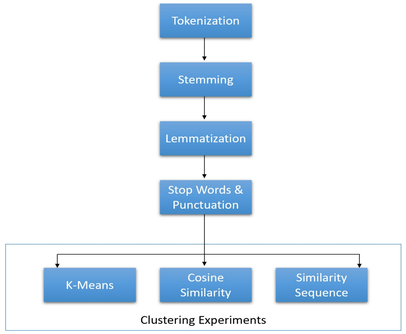
Due to the ever-accelerating development of technological practices, there exists a growing complexity in the meaning of concepts in cybersecurity. This complexity has led to a need for a standardized knowledge base. There are many different existing standards and frameworks that define cybersecurity concepts. We are focused on relating the NICE Framework's tasks, knowledge, and skills (TKS) and the NIST Framework's Controls from Special Publication 800-53. Joining these standards will provide a standard manner for communicating needs for cybersecurity activities in an organization. The challenge presented is to improve upon the (CTB) CyberTalent Bridge tool's matching algorithm for mapping the NIST Controls to the NICE TKS, resulting in the most well-skilled available talent.

Method

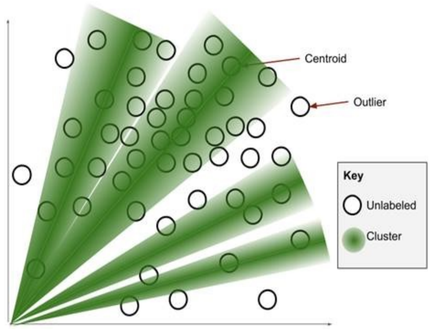
To solve the above challenge, we evaluated the mapping of the NIST Controls and the NICE TKS taking place on the CTB website. Using Python programming we executed text matching to classify the descriptions of the TKS from the NICE 'Reference Spreadsheet for the Workforce Framework for Cybersecurity' to the descriptions of the NIST Controls. By using the NLTK library in Python we implemented word vectorization, tokenized the dataset, and later manipulated the stop words. First, we executed K-Means clustering algorithm that would cluster TKS and group them with the most related Control. After experimenting with different cluster sizes, we concluded that the findings were inconsistent. Next, we looked at text matching methods, such as the similarity sequence matcher and cosine similarity clustering.

Method Explained

Text Vectorization



Cosine Similarity Clustering



Results

Control	TKS	Top 10 Matching TKS	Cosine Similarity	Assessment of Relevance (0-5)
IR-4 INCIDENT RESPONSE	Knowledge	K0042	0.519	5
		K0041	0.350	5
		K0150	0.307	5
		K0230	0.257	5
		K0292	0.219	5
		K0368	0.175	5
		K0481	0.123	5
		K0572	0.110	5
		K0399	0.110	5
		K0222	0.105	5
	Skill	S0054	0.415	5
		S0365	0.307	5
		S0176	0.180	5
		S0350	0.175	5
		S0337	0.154	4
		S0150	0.123	5
		S0200	0.123	4
		S0032	0.116	5
		S0100	0.110	5
		S0309	0.110	5

Conclusions

As of recent, after implementing text mining, word vectorization, text matching, and clustering techniques, we have concluded that clustering TKS is the right approach to mapping the NIST Framework's Controls to the NICE Framework's Task, Knowledge, and Skills. After testing three clustering techniques and discovering that the results of K-Means clustering were unreliable, due to a found inconsistency in varying runs. We have concluded that the cosine similarity algorithm has yielded stable results after repeated experiments. While this method has resulted in relevant findings after again being manually examined, the results are promising, the method is to be further examined and tested.

References

Description	Reference Link
CyberTalent Bridge Tool	https://cybertalentbridge.com/
NIST Special Publication 800-181 R1 (TKS)	https://www.nist.gov/itl/applied-cybersecurity/nice/nice-framework-resource-center/nice-framework-supplemental-material
NIST Special Publication 800-53 R5 (Controls)	https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf
Clustering in Python	https://towardsdatascience.com/clustering-documents-with-python-97314ad6a78d
Cosine Similarity Clustering	https://www.machinelearningplus.com/nlp/cosine-similarity/
Text Mining	https://towardsai.net/p/data-mining/text-mining-in-python-steps-and-examples-78b3f8fd913b
Text Vectorization	https://towardsdatascience.com/understanding-nlp-word-embeddings-text-vectorization-1a23744f7223

Acknowledgements

This research was performed under an appointment to the U.S. Department of Homeland Security (DHS) Science & Technology (S&T) Directorate Office of University Programs Summer Research Team Program for Minority Serving Institutions, administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and DHS. ORISE is managed by ORAU under DOE contract number DE-SC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of DHS, DOE or ORAU/ORISE.

ILLINOIS