Named Entity Recognition and Information Extraction for Deep Web / Dark Web Forum Messages

Abubakar Ahmad Abdullahi
Computer Engineering Department
Marmara University
Istanbul, Turkey
abubabakr@gmail.com

Abstract

The deep web consists of web pages that search engines cannot index. The dark web is a subsection of the deep web which includes illicit and criminal activities. Conversations in dark web forums contain information that may assist in uncovering criminal acts. In this research, I plan to predict the leadership and potential organizational structure of dark web forums using Natural Language Processing. Much of the data for this research will be based on datasets provided by Arizona State University's Artificial Intelligence Laboratory.

Project Schedule

Activity	**Description**	**Duration**
Research and Analysis	Get enough background knowledge to start implementing the project	5 days **(16th - 20th November)**
Algorithms Design	Design algorithms based on the research done	5 days **(21st - 25th November)**
Development: First Phase	Implement algorithms. This step is important to comprehend misunderstood information	10 days **(25th November - 4th December)**
Initial Evaluation and Plan Review	Evaluate initial results from the previous step and plan further improvements	15 days **(4th - 18th December)**
Report Update	Update the research report based on new findings	5 days **(18th - 23rd December)**
Development: Second Phase	Implement new information gathered from the previous steps	15 days **(23rd December - 6th January)**

Final Evaluation and Conclusion	Analyse findings to make informed conclusions	5 days **(6th January - 10th January)**
Final Report	Complete research report	10 days **(10th January - 20th January)**

References

1. Arizona State University's Artificial Intelligence Laboratory (https://www.azsecure-data.org/dark-web-forums.html)
2. Cyber Threat Discovery from Dark Web by Azene Zenebe, Mufaro Shumba, Andrei Carillo and Sofia Cuenca. EPiC Series in Computing Volume 64, 2019, Pages 174–183 Proceedings of 28th International Conference on Software Engineering and Data Engineering
3. Prof. Tunnard's Social Network Analysis and Social Media Classes (http://crtunnard.blogspot.com/2017/10/a-light-in-dark-using-social-network.html)
4. I. Deliu, C. Leichter and K. Franke, "Extracting cyber threat intelligence from hacker forums: Support vector machines versus convolutional neural networks," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 3648-3656, doi: 10.1109/BigData.2017.8258359.