

Thesis Portfolio

Title of Technical Report

Developing Spam Detection and Prevention Schemes using Natural Language Processing

Title of STS Research Paper

An Analysis into the Efficiency of Makerspaces

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering and Applied Science

Parv Ahuja

Spring, 2020

Table of Contents

Sociotechnical Thesis

Technical Thesis: Developing Spam Detection and Prevention Schemes using Natural Language Processing

STS Research Paper: An Analysis into the Efficiency of Makerspaces

Sociotechnical Thesis

Technology's prevalence in society today relies as much on hype as it does on utility, with certain innovations gaining and losing traction solely based on the social perspective of that technology. My theses are related by looking at two different realizations of social hype within technology: Machine Learning and Makerspaces. Both of these technologies have been put on a pedestal as unique ways to solve common problems, but my approach to understanding both differ. The technical thesis involves an in-depth application of Natural Language Processing to solve the novel problem of detecting and preventing spam, while the STS research focuses on specifically separating hype from fact by providing an unbiased way to measure the efficacy of Makerspaces as an educational tool. This report looks at both of these technologies separately, with the technical thesis applying the technology, and the STS research dissecting the technology.

The technical thesis focuses on two aspects of detecting and preventing spam using a novel Machine Learning technique referred to as Natural Language Processing. Initially, work was done on a chat bot placed in the undermarket web to communicate with e-commerce miscreants and extract useful intel. For this project, Natural Language Processing was used to improve the machine's understanding of the conversation with the miscreant, thus enabling it with richer information. Utilizing a new NLP framework, we were able to substantially improve the understanding of textual information and flow of conversation for the chat bot. Additionally, in a separate project, NLP is being used to build a developer's reputation and predict whether

that user is committing malicious code to GitHub. We are currently extracting textual information from committed code and vectorizing to produce more features for the prediction.

The STS thesis investigates the efficiency of Makerspaces as a learning tool, attempting to separate what is fact and what is hype. As many schools begin to adopt Makerspaces, the need to define the metrics by which to evaluate these collaborative spaces increases. In this research, I discuss the existing methods of evaluation from both an educational and social lens. Furthermore, I produce metrics that additionally account for the resources used, the opportunity cost of those resources, and the specific benefits reaped. Through case studies of existing Makerspaces, it was found that successful makerspaces were coupled with an existing curriculum and had strong support from patrons. The research recommends an amended evaluation metric for Makerspaces, as well as key factors that contribute to the efficiency of these spaces. The research is preliminary yet substantive and further research into Makerspaces could be grounded in what is found here.