Thesis Project Portfolio

Using Machine Learning to Detect Plagiarism in Written Works

(Technical Report)

Minimizing Gentrification in Tech Hubs

(STS Research Paper)

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

Wonyoung Choi

Spring, 2022

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

As technology and society continue to mutually advance, the study of how the two are interconnected becomes an integral point of discussion. The growth of plagiarism and gentrification in tech hubs are keen examples of how society and technology are intertwined. The growth of the internet results in larger written sources which can contribute to plagiarism which has its own social implications and motivations. Also, as technology companies become more profitable, gentrification in tech hubs continues to grow resulting in changes to society in those locations. My portfolio covers both of these topics which includes a technical paper addressing a potential solution to plagiarism by creating a machine learning tool for plagiarism detection and an STS paper analyzing the implication of gentrification in large tech hubs.

For my technical paper, I will be explaining a group project I worked on to create a program that detects instances of plagiarism in written works using machine learning. This project was inspired while I was taking a machine learning class at the University of Virginia which sparked the idea of applying what I had learned to address an issue commonly faced in academic institutions. Creating original works is a component of education that is very highly emphasized in the University of Virginia to the point where there is an honor code which must be pledged by students on originality upon submitting their own written works. Despite this, plagiarism is still an issue that remains at the University and other academic institutions. This was the basis of my inspiration to create a machine learning application that would process large amounts of data composed of written works found on the internet. The application would then develop a machine learning model that can determine the similarity between the submitted written work and that of any other piece of literature found in the dataset. The application works as intended and can be used as a basis for future projects that may expand on the idea.

My STS paper addresses the issue of gentrification in tech hubs. I was inspired to write about this issue because I was always aware of gentrification happening in areas such as San Francisco and how large technology companies played a large role. Being a resident of Northern Virginia, I worry that gentrification is already occurring in my home town especially as Amazon continues to build their new headquarters in Arlington, Virginia. I think completely understanding the different actors in gentrification is integral to fully analyzing the situation and producing potential solutions to minimize gentrification. In the paper, I will be using Actor Network Theory to identify the different actors in gentrifying tech hubs and analyzing the motives and needs of each group and how they affect gentrification.

While working on my technical project, I realized that there is a larger societal aspect to plagiarism which is interconnected to technology. Plagiarism becomes a more potent issue as technological advances in the internet are made and in return, new technologies can be created to push back against the growth of plagiarism. This can be compared to my STS paper which addresses how the rise in technology based corporations is fueling gentrification creating societal circumstances that must be addressed in relation to technology.