On the Design of a Database for the Linguistic Atlas Project

Simon Harris-Palmer

We discuss the initial solution for a database support for the Linguistic Atlas Project (LAP). The Linguistic Atlas Project (LAP), founded in 1929 at the behest of the American Dialect Society, is the most thorough and expansive study of American English undertaken to date. Despite this fantastic feat, most of the vast amount of data remains inaccessible to researchers and other data users in a convenient way. The data remains stored in a mix of CSV, Excel, and image files that are hard to navigate and have inconsistencies in how information is recorded. With better access, the door would open for the LAP to be used as a rich pool of data in broader research into American English dialects. The goal is to provide an efficient system to respond to complex queries in the context of the American dialects.

 To complete this goal, I am using the programming language Python 3, along with some utility modules, such as Flask, to create a website and WTForms for handling complex online forms and queries. I am also using SQLite, which allows me to construct and query databases. Tying all of these components with HTML, CSS, and JavaScript for added functionality and visual appeal will produce a useful and user-friendly searching system.

This project is a part of the KY Young Researchers program with Drs. Allison P. Burkette and Jerzy W. Jaromczyk as my research mentors. This research project is being conducted throughout my senior year of high school, having started the year prior.