About Job 02 : Using Web Crawler to Correct Address

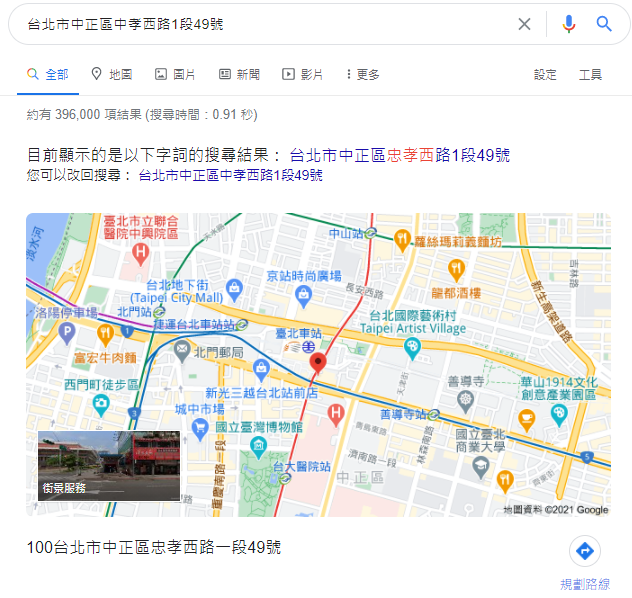
The most important problem which data analyst faced is raw data is quite dirty. When I want to do geography analysis, geography raw data must be transformed into completed addresses or latitude-longitude data.

Take Taipei Station for example:

“台北市中正區忠孝西路1段49號” must be transformed into “臺北市中正區黎明里3鄰忠孝西路一段49號”

TGOS is a useful website where Taiwanese government offer users batch comparison of addresses. However, there is a big problem that TGOS can’t transform address including wrong letters such as “台北市中正區中孝西路1段49號”.

To solve this problem, I found that when I put the wrong address to google and google sometimes will feedback the right address.



Thus, I used python to automate correct addresses. Although it can not 100% solve the problem and cost much time, we can improve the rate of comparison at TGOS.

It’s an article that I want to record my works and practice my English writing. If there is any suggestion or critique, please don't hesitate to give me any comment. I also upload python code and sample data to github. Website: <https://github.com/GallonShih/Projects/tree/main/AboutJob/ReviseAddress>