

Big data techniques for supplementing statistical business registers

Tomasz Klimanek, Statistical Office in Poznań, t.klimanek@stat.gov.pl

Karol Potocki, Statistical Office in Poznań, k.potocki@stat.gov.pl

Webscraping techniques are regarded as an increasingly promising source of data for official statistics. Among many benefits of collecting and processing scraped data, one can mention the possibility of updating information in existing registers maintained by public administration agencies, such as the business register (based on the internet address, social media activity, e-commerce and others).

The authors want to share their experiences of developing methods of collecting and analysing Internet data about companies in Poland.

They are going then to present the initial results of using pattern matching on extracted data from webpages to find some basic contact information which is missing in statistical official business register (eg. telephone number, e-mail address, tax id number, company registration number, etc.)