WHOIS data opening up a new avenue for the measurement of entrepreneurship

A WhoisXML API Customer Success Story

October 2017.

Entrepreneurship is the central process through which economic growth and performance is fostered in a regional economy. Its evaluation is thus of paramount importance for policymakers and economists. However, the quantification of entrepreneurship, that is, introduction of measures to describe the attempts to start growth-oriented business, the likelihood of their success, the ability to raise venture capital, etc. is a challenging task due to the lack of globally available and accurate input data on e.g. business registration.

In a recent working paper a new approach is proposed to overcome this issue by using WHOIS registration data [1]. The new approach is applied to companies in Oxford, UK and Cambridge, UK as a demonstration, by using data purchased from WhoisXML API.

The researchers

NESTA is a global innovation foundation based in the UK. They work with partners from all over the world, ranging from national governments and global technology firms to community groups and local activists. They foster new ideas in order to tackle big global challenges, include the pressures of an aging population; stretched public services; and a fast changing jobs market [2]

Here we summarize briefly some of the ideas in a recent working paper published by NESTA [1], by Abhishek Nagaraj, UC Berkeley and NBER Sibo Wang, Massachusetts Institute of Technology.

WHOIS-based entrepreneurship quantification

There has been a significant effort aimed at the proper quantifications of entrepreneurship. There are very powerful approaches available, however, they have to rely on sources of business registration data. These data come from sources which are typically limited to areas or regions, have a limited time range. In addition, they are frequently incomplete or simply unavailable.

The new methodology in Ref. [1] is built on the observation that most growth-oriented businesses begin not with company registration, but with domainname registration. This is often carried out by entrepreneurs before starting the actual business, thus domain name registration data can harbor useful information predicting actual entrepreneurship activity. WHOIS data are reach and sometimes provide a direct indicator of the entrepreneurial intent of the company, including information on the sector of the business and the orientation of growth.

The researchers have purchased WHOIS data from WhoisXML API, as these are collected systematically, they are unified and consistent, and they are provided in formats ready to use in research. These data served as the solid foundation of the research. They have supplemented these data with information from other sources. For instance, to identify entrepreneurial activity the contents of actual web pages have to be analyzed, too. A web page may contain "bad keywords" indicating that it is a blog or some personal information, which is not relevant for the present analysis. The researchers have collected such data by scraping, and they have used archive web data, too. These were processed by text-analysis techniques. As WHOIS data also contain addresses of registrars, they can be supplemented with geographical information resulting in excellent granularity at a micro-geographic level. Thanks to the availability of WHOIS data on a long timespan and the existence of archived web pages, the level of granularity in time is also outstanding. The collected data were analyzed by advanced machine learning techniques.

Conclusions

By the detailed analysis of entrepreneurship in Oxford, UK and Cambridge, UK, in Ref. [1] it has been demonstrated that the analysis of domain registration data opens new perspectives in understanding entrepreneurial activity. Relying on global, fundamentally open and available data, this approach can be used to measure this kind of activity across many regions of the world, and provides informa-

tion at an extremely granular level both over time and microgeographically.

References

[1] Abhishek Nagaraj and Sibo Wang. Quantifying entrepreneurship using domain name registration data: Methods and applications for Oxford and Cambridge, UK. *NESTA Working Paper 16/02*. September 2016.

[2] NESTA. `http://www.nesta.org.uk`, visited on 2017.10.11.