Using Entrepreneurs' Digital Footprint for Analyzing a Country's Entrepreneurial Activity

Andrei Parfenov[1], Zeljko Tekic[2] and Maksim Malyy[3]

[1]Practicum Tech, Belgrade, Serbia
[2]Graduate School of Business, HSE University, Moscow, Russian Federation
[3]KEC – Kreativno Edukativni Centar, Novi Sad, Serbia

Introduction

The wide usage of search engines led to the growth of a digital footprint that people leave on the Web. Some search engines (e.g., Google, Yandex, Baidu) provide tools for exploring search statistics on certain queries. Search query data have been proved to be very useful in predicting some number off underlying social and economic trends (Choi and Varian, 2009; Jun et al., 2018; Wu and Brynjolfsson, 2009). Building on this evidence, we examine hypothesis that Internet search query data on specific terms related to opening/registering SMEs correlate and match dynamics of real-life SME registrations in a country. If the hypnotized relationship exists, it may serve as a bases for an alternative methodology for analyzing and forecasting entrepreneurial activity in a country, which is important as SMEs contribute significantly to national income in all economies, while assessing the effectiveness of government support measures for SMEs can be difficult (OECD, 2021). Especially in countries with a lack of transparency, where governments are policymakers and data owners.

Data and Methodology

The hypothesis is examined on the example of the Russian Federation. This is one of rare countries where Google is not solely dominant search engine, but it is equally used with the local search engine (i.e., Yandex). Therefore, we use search query data from both engines. On the other hand, we use monthly statistics on newly registered SMEs in the country from an open database provided by Federal Tax Service in Russia (Federal Tax Service, 2021).

To understand what founders of new businesses are searching for when they consult the Internet for help, we interviewed 10 owners of recently established businesses in Russia. Based on the interviews, eight specific queries related to the process were identified and associated in three groups (Table 1). Keywords from Group 1 are related exclusively to opening a sole proprietorship,

from Group 2 to opening a limited liability company, while those from Group 3 are related to opening both – a sole proprietorship and limited liability company.

Table 1. List of keywords used for search identified during the interviews

Group	English (translation)	Russian (original)
1	sole proprietorship	индивидуальный предприниматель (ИП)
	registration of sole proprietorship	регистрация ИП
	opening sole proprietorship	открыть ИП
2	limited liability company (LLC)	общество с ограниченной ответственностью (ООО)
	registration of LLC	регистрация ООО
	opening of LLC	открыть ООО
3	business idea	бизнес-идея
	charter documents	учредительные документы

Data on both types of SMEs – sole proprietorships and limited liability companies (LLCs) – registered in 2018-2021 are derived from the open database of Federal Tax Service for every month and collected in a time series. Search query statistics on related terms were obtained for the same period from Google Trends (GT). Time series from Yandex Wordstat (YW) are obtained for the period of 2020-2021 due to the tool's limitation (i.e., statistics are only available for the last two years). YW provides precise search data for all the regions of Russia, while GT provides data with sudden outbreaks and fast drops to zero for most of the regions. Thus, GT data were retrieved for the Russian Federation overall, not for its regions.

Correlational analysis was performed for individual regions and Russia overall. Since linear relationship between datasets is unclear and the size of the datasets is small, the rank correlation coefficient Kendall's tau is chosen to provide the most reliable results (Akoglu, 2018). Based on Kendall's tau (τ) connection with Pearson's r (Walker, 2003) $r = \sin(0.5*\pi*\tau)$, and following Malyy et al. (2021) we selected 0.5, for the threshold level, thus, treating the equal to and higher values as evidence of a strong link (the moderate link is 0.3–0.49, the weak if less than 0.3).

Results

Our results show that aggregated and anonymized data about web searches of millions of people, provided by relevant big data tools, hold a promise to help us understand the entrepreneurial intentions and actions of a large population well without conducting costly and time-consuming surveys. Usage of the Internet search query data for this purpose is more robust when: 1) procedures for registering a business are simpler, especially if everything could be done online; 2)

the unit of analysis (a country or a region) has more digitally developed support services; and 3) significant change (e.g., policy measure) is introduced in the system (a country or a region). In this way, we contribute to the ongoing debate about the value of big data and the Internet traces for entrepreneurship research (Malyy et al., 2021; Prüfer and Prüfer, 2020; Schwab and Zhang, 2019). Our results may serve as a starting point for a new approach to measuring entrepreneurial activity in a country or a region, for understanding the dynamics of new business formation, and for predicting their number. Thus, the results offer clear implications for addressing policymaking issues related to entrepreneurial activities in a country (e.g., monitoring the effects of creating infrastructure, or implementing special programs and incentives).

References

Akoglu, H., 2018. User's guide to correlation coefficients. Turkish J. Emerg. Med. 18, 91–93.

Choi, H., Varian, H., 2009. Predicting the present with Google Trends, Google Inc.

Jun, S.P., Yoo, H.S., Choi, S., 2018. Ten years of research change using Google Trends: from the perspective of big data utilizations and applications. Technol. Forecast. Soc. Change 130, 69–87.

Malyy, M., Tekic, Z., Podladchikova, T., 2021. The value of big data for analyzing growth dynamics of technology-based new ventures. Technol. Forecast. Soc. Change 169, 120794.

OECD, 2021. OECD SME and Entrepreneurship Outlook 2021, OECD SME and Entrepreneurship Outlook 2021. OECD.

Prüfer, J., Prüfer, P., 2020. Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in the Netherlands. Small Bus. Econ. 55, 651–672.

Schwab, A., Zhang, Z., 2019. A New Methodological Frontier in Entrepreneurship Research: Big Data Studies. Entrep. Theory Pract. 43, 843–854.

Walker, D.A., 2003. Converting Kendall's Tau For Correlational Or Meta-Analytic Analyses. J. Mod. Appl. Stat. Methods 2.

Wu, L., Brynjolfsson, E., 2009. The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. SSRN Electron. J.